Distributed Functional Scalar Quantization: High-Resolution Analysis and Extensions

Vinith Misra, Vivek K Goyal, *Senior Member, IEEE*, and Lav R. Varshney, *Graduate Student Member, IEEE*

Abstract

Communication of quantized information is frequently followed by a computation. We consider situations of *distributed functional quantization*: distributed quantization of (possibly correlated) sources followed by centralized computation of a function. Under smoothness conditions on the sources and function, asymptotically-optimal regular scalar quantizer designs are developed to minimize distortion of the computed function. Striking improvements over quantizers designed without consideration of the function are possible and are larger in the entropy-constrained setting than in the fixed-rate setting. As extensions to the basic analysis, we characterize a large class of functions for which regular quantization suffices, consider certain functions for which asymptotic optimality is achieved without arbitrarily fine quantization, and allow limited collaboration between source encoders. In the entropy-constrained setting, a single bit communicated between encoders can have an arbitrarily-large effect on functional distortion. In contrast, such communication has very little effect in the fixed-rate setting.

Index Terms

Asymptotic quantization theory, distributed source coding, non-difference distortion measures, optimal point density function, rate-distortion theory

I. Introduction

CONSIDER a collection of spatially-separated sensors, each measuring a scalar X_j, $j = 1, 2, \ldots, n$. As shown in Fig. 1, the measurements are encoded and communicated over rate-limited links to a sink node without any interaction between the sensors. The sink node computes an estimate of the function $g(X_1^n) = g(X_1, X_2, \ldots, X_n)$ from the received data. Coding that exploits statistical dependence among the X_js is commonly called *distributed source coding* and has been the subject of much research. A complementary concept is to exploit the form of the function g in designing the (separate) coding of each measurement. Restricting to scalar quantization, this *distributed functional scalar quantization* (DFSQ) problem is the central subject of this paper. Optimal DFSQ can provide performance improvements in addition to any that are rooted in statistical dependence of the X_js; thus for clarity, most examples presented here are for cases with independent X_js.

The term *functional source coding* (FSC) can be reasonably applied any time the information sink uses an approximation to the evaluated function $g(X_1^n)$ instead of the source variables X_1^n directly. For emphasis, we will refer to approximate representation of X_1^n as *ordinary* source coding. FSC is a trivial problem when the encoding is centralized; in that case the encoder mapping can be the composition of the function g and a good encoder for the random variable $g(X_1^n)$. With the constraint of distributed encoding, no single encoder can compute the function, and the situation is thus more intricate.

The primary aim of this paper is to develop a high-resolution approach to optimal DFSQ. Here as in ordinary source coding, the high-resolution approach yields optimality among regular quantizers. In ordinary source coding, this is an insignificant limitation because, quite generally, optimal quantizers are regular. For DFSQ, some restrictions on g are needed to ensure that the optimal quantizers are regular. This provides another key contrast to previous work. Using only the graph coloring approach to FSC of Doshi *et al.* [1] provides no improvement under these restrictions on g, so the present work is a complement to [1]. Combining the two approaches is discussed in Section VII.

A. Basic Problem Statement

An information sink wishes to obtain an estimate of $g(X_1^n)$ where $g : \mathbb{R}^n \to \mathbb{R}$ satisfies some smoothness conditions and the random variables $\{X_j\}_{j=1}^n$ (denoted more compactly X_1^n) have some known joint distribution. The estimate $\widehat{g}(\widehat{X}_1^n)$ is computed from scalar-quantized values

$$\widehat{X}_j = Q_j(X_j), \qquad j = 1, 2, \ldots, n,$$

where Q_j applied to X_j has rate R_j. In the fixed-rate (codebook-constrained) setting, this means Q_j has $K_j = 2^{R_j}$ levels; in the variable-rate (entropy-constrained) setting, this means $H(Q_j(X_j)) = R_j$ where $H(\cdot)$ denotes the entropy.

This work was supported in part by NSF Grant CCF-0729069.

The material in this paper was presented in part at the Information Theory and its Applications Workshop, La Jolla, California, January/February 2008; and the IEEE Data Compression Conference, Snowbird, Utah, March 2008.

V. Misra (email: vinith@stanford.edu) was with the Massachusetts Institute of Technology when this work was completed and is now with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA. V. K. Goyal (email: vgoyal@mit.edu), and L. R. Varshney (email: lrv@mit.edu) are with the Department of Electrical Engineering and Computer Science and the Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139 USA. L. R. Varshney is also with the Laboratory for Information and Decision Systems.

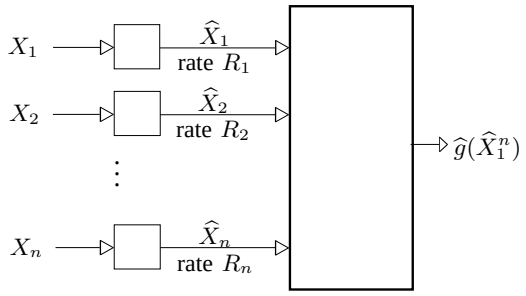


Fig. 1. Distributed functional source coding.

The accuracy of the approximation is measured by the mean-squared error (MSE)

$$D = \mathbf{E}\left[\left(g(X_1^n) - \widehat{g}(\widehat{X}_1^n)\right)^2\right].$$

For a given set of rates $\{R_j\}_{j=1}^n$ or a maximum sum rate, we seek designs of the quantizers $\{Q_j\}_{j=1}^n$ such that distortion D is minimized. The problem is approached under the standard assumptions for high-resolution analysis [2], and restrictions on g are applied as needed.

B. Ramifications and Extensions

Unsurprisingly, there are situations in which designing quantizers to minimize D is no different than designing them for low MSEs $\mathbf{E}[(X_j - \widehat{X}_j)^2]$, $j = 1, 2, \ldots, n$. Our analysis will show, for example, that there is no advantage from accounting for g when g is linear. However, there are also cases in which the improvement is very large for large values of n; examples in Section V display distortion improvement over ordinary source coding by a factor that is polynomial in n in the fixed-rate case and exponential in n in the variable-rate case.

In addition to the basic formulation, we consider extensions that re-examine requirements on g and on the lack of communication among encoders. First, we define a requirement termed *equivalence-free* that is less restrictive than monotonicity but still guarantees that optimal quantizers are regular at sufficiently high rate. This leads also to some consideration of non-regular quantizers. Next, we explore a situation in which the high-resolution analysis breaks down because there is an interval where the marginal density f_{X_j} is positive but the optimal quantizer for X_j seems to not have fine partitions. This prompts the concept of a *don't care interval*, a mixture of low- and high-resolution, and connections with [1]. Finally, we allow rate-constrained information $Y_{2\to1}$ communicated from encoder 2 to encoder 1 to affect the encoding of X_1. We call this *chatting* and bound its effect on the distortion D. In the fixed-rate setting, the reduction in distortion from $Y_{2\to1}$ can be no more than if R_1 were increased by the same rate; in the variable-rate setting, the reduction in distortion can be arbitrarily large.

C. Structure of Paper

We start in Section II by discussing several topics with connections to functional quantization. Additionally, we briefly review the high-resolution approximation techniques used in our analysis. In Section III we obtain optimal fixed- and variable-rate functional quantizers for the $n = 1$ case; while not important in practice, this case illustrates the role of monotonicity and smoothness of $g(\cdot)$. Generalizations to arbitrary n, under monotonicity restrictions on $g(\cdot)$, are given in Section IV. Some notable examples in Section V are those that show dramatic scaling of distortion with respect to n.

The second half of the paper extends the basic theory of Section IV. Section VI addresses the monotonicity restriction and shows that a weaker *equivalence-free* condition is sufficient for the optimality of the constructions of Section IV to hold. In the process we develop the notion of high-resolution non-regular quantization. In Section VII, we consider certain conditions that cause the high-resolution approach to lead to an optimal quantizer for X_j that does not have high resolution over the entire support of f_{X_j}. A modified analysis and design procedure yields a "rate amplification" in the variable-rate case. Limited communication between encoders, or chatting, is studied in Section VIII, concluding comments appear in Section IX.

II. Background

A. Related Work

DFSQ lies at the intersection of several problems including quantization, distributed source coding, and non-MSE distortion measures. As such, there are many connections to related work. We provide a brief summary of some of these connections here.

Consider the situation depicted in Fig. 1 with $n = 2$. In general, X_1 and X_2 are random variables with some joint distribution, and g is a function of the two. We arrive at several related topics by considering special cases of this formulation.

- If g is the identity function, we have a general distributed source coding problem that is well-known in the lossless setting [3]. The lossy setting with scalar coding is considered in [4], and the lossy problem with jointly Gaussian sources and MSE distortion was recently solved in [5]. In this situation, the correlation of X_1 and X_2 is of primary interest.
- In the lossless setting, Han and Kobayashi [6] studied the classification of functions according to whether the rate region is the same as that for the identity function (i.e., the same as the Slepian–Wolf rate region). Their results are conclusive when $n = 2$ and the source alphabets are finite.
- If $g(X_1, X_2) = X_1$ and R_2 is unconstrained, then X_2 can be viewed as receiver side information available at the decoder. The trade-off between R_1 and distortion (of X_1 alone) is given by the Wyner-Ziv rate-distortion function [7], [8]. Rebollo-Monedero *et al.* examined the Wyner-Ziv scenario at high resolution and showed that providing the receiver side information to the encoder yields no improvement in performance [9].
- For general g and R_2 unconstrained, the problem has been studied by Feng *et al.* [10], who provide an assortment of rate-loss bounds on performance.
- Under suitable constraints on the distortion metric, one may also view X_2 as receiver side information that determines the distortion measure on X_1, drawing a connection to [11] and [12].
- For discrete X_1 and X_2, the lossless regime has been explored for R_2 unconstrained by Orlitsky and Roche [13]. The distributed version of this problem, which involves minimizing the sum-rate $R_1 + R_2$, was later explored by Doshi *et al.* [14].
- Let $Y = g(X_1, X_2)$. Then Y may be interpreted as a *remote source* that is observed only through X_1 and X_2, and we have the remote source multiterminal source coding problem [15].

Interesting related problems have also arisen without a requirement of distributed coding. Rather than having a single function g, one may consider a set of functions $\{g_a\}_{a\in\mathcal{A}}$ and define

$$D_g = \mathbf{E}\left[d(g_\alpha(X_1^N), g_\alpha(\widehat{X}_1^N)\right],$$

where α is a random variable taking values in index set $\mathcal{A}$. One may consider this a special case of the Wyner-Ziv problem with α as decoder side information and a functional distortion measure. In such a setting, fixed- and variable-rate quantization to minimize MSE was studied by Bucklew [16]. Note that if the function were known deterministically to the encoder, one could do no better than to simply compute the function and encode the result.

Under appropriate constraints on the function g, one may consider it as having introduced a *locally quadratic* distortion measure on the source X_1^n. In [17], Linder *et al.* consider quantization via companding functions for locally quadratic distortion measures. We say more about connections to this work in Section IV-E.

Additionally, quantization with a functional motive bears resemblance to the idea of "task-oriented quantization." There has been considerable work in this direction for detection [18], [19], classification [20], and estimation [21]; see also the review article [22]. The use of a function at the decoder can be seen as inducing a non-MSE distortion measure on the source data. In this sense, a thread may be drawn to perceptual source coding [23], where a non-MSE distortion reflects human sensitivity to audio or video.

B. High-Resolution Approach to Quantizer Design

We first provide an informal summary of the assumptions and approximations that are standard for high-resolution analyses. Then, optimization of quantizers under these high-resolution approximations are summarized. More technical details and references to original sources may be found in [2].

1) Assumptions and Basic Approximations: Let X be a random variable with probability density function (pdf) $f_X(x)$. Suppose a quantizer for X has points $\{\beta_i\}_{i\in\mathcal{I}}$ and partition $\{S_i\}_{i\in\mathcal{I}}$. For optimality, it is necessary for each set in the partition to be an interval, i.e., the quantizer is *regular* [24, Sect. 6.2].

The distortion of the quantizer is

$$\begin{aligned} D_X &= \mathbf{E}\left[(X-\widehat{X})^2\right] \\ &= \sum_{i\in\mathcal{I}} \mathbf{E}\left[(X-\beta_i)^2 \mid X\in S_i\right]\mathbf{P}\left(X\in S_i\right) \end{aligned} \tag{1}$$

by the law of total expectation. The initial aim of high-resolution theory is to express this distortion as an integral involving f_X. To that end, we make the following assumptions about the source and quantizer:

HR1. f_X is smooth enough that it may be approximated as constant on each S_i. While it is convenient to think of f_X as continuous, it suffices for it to be measurable [2, Sect. IV-A].

HR2. f_X has bounded support or decays sufficiently fast. Sufficient decay is for terms in (1) corresponding to unbounded S_is to make negligible contributions.

HR3. Neighboring cells have approximately equal sizes, except possibly for two semi-infinite boundary cells.

When the number of points $K = |\mathcal{I}|$ is large, Assumption HR3 allows one to define a (normalized) *point density function* $\lambda(x)$ such that $\delta\lambda(x)$ is approximately the fraction of quantizer points in an interval of length δ centered at x. The point density is used to express the lengths of the partition cells:

$$x \in S_i \quad \Rightarrow \quad \text{length}(S_i) \approx (K\lambda(x))^{-1}. \tag{2}$$

Now we can approximate each non-boundary term in (1). By Assumption HR1, β_i should be approximately at the center of S_i, and the length of S_i then makes the conditional expectation approximately $\frac{1}{12}(K\lambda(\beta_i))^{-2}$. Invoking Assumption HR1 again, the ith term in the sum is $\int_{x\in S_i} \frac{1}{12}(K\lambda(\beta_i))^{-2} f_X(x)\,dx$. Finally, neglecting overload distortion because of Assumption HR2,

$$D_X \approx \int \frac{(K\lambda(x))^{-2}}{12} f_X(x)\,dx = \frac{1}{12K^2}\mathbf{E}\left[\lambda^{-2}(X)\right]. \tag{3}$$

This approximation holds in the sense that the ratio of the two quantities approaches 1 as the rate increases. This is the meaning of "$\approx$" for the remainder of the paper, except where noted.

In general, the optimal variable-rate quantizer may have an infinite number of points, and this is handled with an unnormalized point density. For convenience, we consider sources with support bounded to $[0, 1]$ to obviate this. Then as long as the quantization is fine ($\lambda(x) > 0$) wherever the density is positive, we can approximate the output entropy of the quantizer using the point density as follows:

$$\begin{aligned}
H(\widehat{X}) &= -\sum_{i\in\mathcal{I}} \mathbf{P}\left(X \in S_i\right) \log_2 \mathbf{P}\left(X \in S_i\right) \\
&\overset{(a)}{\approx} -\int f_X(x) \log_2 p(x)\,dx \\
&\overset{(b)}{\approx} -\int f_X(x) \log_2 (f_X(x)/(K\lambda(x)))\,dx \\
&= -\int f_X(x) \log_2 f_X(x)\,dx \\
&\qquad + \int f_X(x) \log_2 (K\lambda(x))\,dx \\
&= h(X) + \log_2 K + \mathbf{E}\left[\log_2 \lambda(X)\right],
\end{aligned} \tag{4}$$

where $p(x)$ is defined as $\mathbf{P}\left(X \in S_i\right)$ for $x \in S_i$ and $h(X)$ is the differential entropy of X. Step (a) uses HR1; and step (b) uses HR1 and (2).

2) Optimal Point Densities: Once quantizer performance has been expressed in terms of point densities, optimal designs can be found easily. We derive the optimizing point densities and the resulting distortions because analogous optimizations appear in Sections III and IV.

In the fixed-rate case, the problem is to minimize D_X for a given value of K. (The rate is $R = \log_2 K$.) An application of Hölder's inequality yields

$$\begin{aligned}
\int f_X^{1/3}(x)\,dx &= \int \left(\frac{f_X(x)}{\lambda^2(x)}\right)^{1/3} (\lambda(x))^{2/3}\,dx \\
&\le \left(\int \frac{f_X(x)}{\lambda^2(x)}\,dx\right)^{1/3} \left(\int \lambda(x)\,dx\right)^{2/3} \\
&= \left(\mathbf{E}\left[\lambda^{-2}(X)\right]\right)^{1/3},
\end{aligned}$$

with equality when $f_X(x)/\lambda^2(x)$ is proportional to $\lambda(x)$. Thus, D_X is minimized by

$$\lambda(x) = f_X^{1/3}(x) / \left(\int f_X^{1/3}(t)\,dt\right). \tag{5}$$

The resulting minimal distortion is

$$D_X \approx \frac{1}{12K^2}\left(\int f_X^{1/3}(x)\,dx\right)^3 = \frac{1}{12}\|f_X\|_{1/3} 2^{-2R}, \tag{6}$$

where we have introduced a notation for the $\mathcal{L}^{1/3}$ pseudonorm.

For the variable-rate scenario, the problem is to minimize D_X for a given maximum value of $H(\widehat{X})$. Starting with a rearrangement of (4) and using Jensen's inequality (with the convexity of $-\log_2(\cdot)$),

$$\begin{aligned}
2(H(\widehat{X}) - h(X)) &\approx \mathbf{E}\left[-\log_2(K^{-2}\lambda^{-2}(X))\right] \\
&\ge -\log_2 \mathbf{E}\left[K^{-2}\lambda^{-2}(X)\right] \\
&\approx -\log_2(12D_X).
\end{aligned}$$

This translates to an approximate lower bound on D_X, and the inequality step holds with equality when $\lambda(X)$ is a constant. Thus $\lambda(x) = 1$ is asymptotically optimal, i.e., the quantizer should be uniform.[1] The corresponding minimal distortion is

$$D_X \approx \frac{1}{12} 2^{2h(X)} 2^{-2R}. \tag{7}$$

Note that both optimal point densities are positive on the entire support of f_X. Thus, at high enough resolution, the quantization is fine *pointwise over* X. In the functional settings, this will be used to justify piecewise linear approximation of the function g. Note also that both variable- and fixed-rate quantization have $\sim 2^{-2R}$, or -6 dB/bit, dependence of distortion on rate. This is a common feature of ordinary quantizers, but we demonstrate in Section VII that certain functional scenarios can cause distortion to fall even faster with the rate.

One way to concretely specify a quantizer from a point density is to require

$$\Lambda(\beta_i) = i - \tfrac{1}{2}, \qquad i = 1, 2, \ldots, K,$$

where $\Lambda(x) = \int_0^x \lambda(t)\, dt$ is the "cumulative" point density. However, analysis of quantizers through point densities does not rely on the precise placement of codewords and cell boundaries. Under the assumptions of high-resolution analysis, $o(1/K)$ deviations in the β_is do not affect the distortion. We return to this point in Section III-E to partially generalize the basic analysis to discontinuous functions.

3) Optimal Bit Allocation: As a final preparatory digression, we state the solution to a typical bit allocation problem that arises several times in Section IV.

Lemma 1: Suppose $D = \sum_{j=1}^n c_j 2^{-2R_j}$ for some positive constants $\{c_j\}_{j=1}^n$. Then the minimum of D over the choice of $\{R_j\}_{j=1}^n$ subject to the constraint $\sum_{j=1}^n R_j \le nR$ is attained with

$$R_j = R + \frac{1}{2} \log_2 \frac{c_j}{\left(\prod_{j=1}^n c_j\right)^{1/n}}, \qquad j = 1, 2, \ldots, n,$$

resulting in

$$D = n \left(\prod_{j=1}^n c_j\right)^{1/n} 2^{-2R}.$$

Proof: The result can be shown using the method of Lagrange multipliers. It appeared first in the context of bit allocation in [25]; a full proof appears in [24, Sect. 8.3]. ■

The lemma does not restrict the R_js to be nonnegative or to be integers. Such restrictions are discussed in [26].

III. Univariate Functional Quantization

Let X be a random variable with pdf $f_X(x)$ defined over $[0, 1]$, and let $g : [0, 1] \to \mathbb{R}$ be the function of interest. The source X is quantized at rate R into $\widehat{X} = Q(X)$, and an estimate $\widehat{g}(\widehat{X})$ is formed at the decoder, where $\widehat{g}$ is the estimator function. We wish to design $\widehat{g}$ and $Q(X)$ to minimize the functional distortion, $D = \mathbf{E}[(g(X) - \widehat{g}(\widehat{X}))^2]$.

Since we seek to answer this design question with high-resolution techniques, the function g and the source X must be restricted in a manner similar to Section II-B. For the moment we err on the side of being too strict. Sections VI and VII will significantly loosen these requirements.

We have already assumed f_X has bounded support. Additionally, we require high-resolution assumptions HR1 and HR3 from Section II-B and the following conditions for the univariate function:

UF1. g is monotonic.

UF2. g is continuous on $[0, 1]$; and g' and g'' exist and are uniformly bounded, except possibly at a finite number of points.

A. Sufficiency of $\widehat{g} = g$

Throughout this paper, we assume that $\widehat{g} = g$. In the univariate case, the assumed continuity of g ensures this is without loss of generality.

Lemma 2: Consider a functional quantization problem with source X and function of interest g that is continuous on the closure of the support of X. Given any quantizer and estimator pair $(Q, \widehat{g})$, there exists a quantizer $\widetilde{Q}$ with the same rate as Q such that the pair $(\widetilde{Q}, g)$ has distortion at most equal to that of $(Q, \widehat{g})$.

Proof: We will prove the lemma by picking a pair $(\widetilde{Q}, \widetilde{g})$ with $\widetilde{g} = g$ that minimizes the functional MSE. Decompose Q as $Q = \beta_Q \circ \alpha_Q$ where $\alpha_Q : \mathbb{R} \to \mathcal{I}$ and $\beta_Q : \mathcal{I} \to \mathbb{R}$, and decompose $\widetilde{Q}$ similarly. Picking $\alpha_{\widetilde{Q}} = \alpha_Q$ makes the rates of Q and $\widetilde{Q}$ equal. It remains to pick $\beta_{\widetilde{Q}}$ that when paired with g minimizes functional distortion.

[1] Recall that for the variable-rate case we are assuming f_X is supported on $[0, 1]$. For other bounded supports, the optimal point density would still be a constant, but perhaps different from 1. Unbounded supports require the use of an unnormalized point density.

For the functional MSE to be minimized, it is necessary for the quantization decoder $\beta_{\widetilde{Q}}$ and estimator $\widetilde{g}$ to satisfy

$$\widetilde{g}(\beta_{\widetilde{Q}}(i)) = \mathbf{E}\left[g(X) \mid X \in S_i\right] \qquad \text{for every } i \in \mathcal{I}.$$

Even with no restriction on g and S_i we must have

$$\min_{x \in \overline{S_i}} g(x) \leq \mathbf{E}\left[g(X) \mid X \in S_i\right] \leq \max_{x \in \overline{S_i}} g(x).$$

Now since g is continuous, the intermediate value theorem implies the existence of a value $\beta_{\widetilde{Q}}(i)$ such that $g(\beta_{\widetilde{Q}}(i)) = \mathbf{E}\left[g(X) \mid X \in S_i\right]$. This specifies the desired quantization decoder $\beta_{\widetilde{Q}}$. ■

Note that the proof of Lemma 2 did not require $g(S_i)$ to be an interval, even though optimal quantization of $g(X)$ involves partitioning into intervals. In the natural case that each $g(S_i)$ is an interval, one can require $\beta_{\widetilde{Q}}(i) \in S_i$ for every i, as one would expect from a quantizer.

B. Sufficiency of Regular Quantizers

The following lemma relates monotonicity to regularity of optimal quantizers, thus justifying the introduction of Assumption UF1:

Lemma 3: If g is monotonic, there exists an optimal functional quantizer of X that is regular.

Proof: The optimal functional quantizer in one dimension is induced by the optimal ordinary quantizer for the variable $Y = g(X)$. That is, one may compute the function $g(X)$ and quantize it directly. Since the optimal ordinary quantizer for a real-valued source is regular, the optimal quantizer over Y, denoted by $Q_Y(y)$ and having points $\{\widehat{y}_i\}_{i \in \mathcal{I}}$, is regular.

$Q_Y(y)$ may be implemented by a quantizer for X with cells given by $g^{-1}(Q_Y^{-1}(\widehat{y}_i))$. We know that $Q_Y^{-1}(\widehat{y}_i)$ is an interval since Q_Y is regular. Also, since g is monotonic, the inverse map g^{-1} applied to any interval in the range of g gives an interval. Thus $g^{-1}(Q_Y^{-1}(\widehat{y}_i))$ is an interval, which demonstrates that a regular quantizer in X will be optimal. ■

C. High-Resolution Distortion

Assumption UF2 is introduced so that a piecewise linear approximation of g suffices in estimating the functional distortion of the quantizer. Recalling the notation $\{\beta_i\}_{i \in \mathcal{I}}$ for the quantizer points and $\{S_i\}_{i \in \mathcal{I}}$ for the partition, we will show that

$$g_{\mathrm{PL}}(x) = g(\beta_i) + g'(\beta_i)(x - \beta_i), \quad \text{for } x \in S_i, \quad i \in \mathcal{I}$$

is an adequate approximation of g for our purposes. Excluding partition cells in which $g''(x)$ does not exist, for any $x \in S_i$,

$$|g(x) - g_{\mathrm{PL}}(x)| \leq \tfrac{1}{2} \left(\max_{\xi \in S_i} |g''(\xi)| \right) (\mathrm{length}(S_i))^2 \tag{8}$$

by Taylor's theorem. Then, invoking Assumption UF2 and the fact that $\mathrm{length}(S_i)$ vanishes for all S_is that intersect the support of f_X, we see that g_{PL} is accurate in a precise sense.

The use of g_{PL} prompts us to give a name to the magnitude of the derivative of g. The distortion is then expressed using this function.

Definition 1: The *single-variate functional sensitivity profile* of g is defined as $\gamma(x) = |g'(x)|$.

Theorem 4: Suppose a source $X \in [0,1]$ is quantized with a K-level quantizer with point density $\lambda(x)$. Further suppose that the source, quantizer, and function $g : [0,1] \to \mathbb{R}$ satisfy Assumptions HR1–3 and UF1–2. Then

$$D = \mathbf{E}\left[(g(X) - g(\widehat{X}))^2\right] \approx \frac{1}{12K^2} \mathbf{E}\left[(\gamma(X)/\lambda(X))^2\right]. \tag{9}$$

Proof: See Appendix A. ■

D. Optimal Point Densities

The distortion expression (9) bears strong resemblance to (3), but with the probability density $f_X(x)$ replaced with a *surrogate density* $\gamma^2(x) f_X(x)$. Optimal point densities and the resulting distortions now follow easily.

For fixed-rate coding, we are attempting to minimize the distortion (9) for a given value of K. Following the arguments in Section II-B2, the optimal point density is proportional to the cube root of the surrogate density:

$$\lambda(x) = \frac{\left(\gamma^2(x) f_X(x)\right)^{1/3}}{\int \left(\gamma^2(t) f_X(t)\right)^{1/3} \, dt}. \tag{10}$$

The (asymptotic) optimality of this point density relies on the quantization being fine everywhere f_X is positive. Thus, we must exclude the possibility that $\gamma(x) = 0$ for an interval $x \in (a,b)$ such that $\mathbf{P}\left(X \in (a,b)\right) > 0$ because in this case the

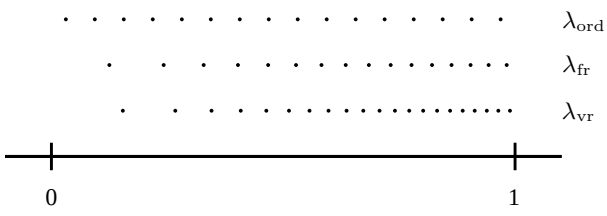


Fig. 2. Quantizer points illustrating the point densities derived in Example 1 at rate $R = 4$.

quantization is not fine for $X \in (a, b)$. We revisit this restriction in Section VII. By evaluating (9) with point density (10), the resulting distortion is

$$D \approx \frac{1}{12} \left\| \gamma^2 f_X \right\|_{1/3} 2^{-2R}. \tag{11}$$

For variable-rate coding, we are attempting to minimize the distortion subject to an upper bound on the rate given by (4). By a derivation similar to that of ordinary variable-rate quantization, the optimal point density is found to be proportional to the functional sensitivity profile:

$$\lambda(x) = \frac{\gamma(x)}{\int \gamma(t)\, dt}. \tag{12}$$

The restriction to make the quantization fine everywhere f_X is positive takes the same form as above. The high-resolution rate approximation (4) is then valid, and the resulting distortion is

$$D \approx \frac{1}{12} \|\gamma\|_1^2 \, 2^{2h(X) + 2\mathbf{E}[\log_2 \gamma(X)]} \, 2^{-2R}. \tag{13}$$

Example 1: Suppose X is uniformly distributed over $[0, 1]$ and $g(x) = x^2$. For both fixed- and variable-rate, the optimal ordinary quantizer is uniform, i.e., $\lambda_{\text{ord}} = 1$. With $\gamma(x) = 2x$, evaluating (9) gives $D_{\text{ord}} \approx \frac{1}{9} 2^{-2R} \approx 0.111 \cdot 2^{-2R}$.

The optimal point density for fixed-rate functional quantization is $\lambda_{\text{fr}}(x) = \frac{5}{3} x^{2/3}$ and yields distortion

$$D_{\text{fr}} \approx \frac{1}{12} \|(2x)^2\|_{1/3} \cdot 2^{-2R} = \frac{9}{125} 2^{-2R} \approx 0.072 \cdot 2^{-2R}.$$

The optimal point density for variable-rate functional quantization is $\lambda_{\text{vr}}(x) = 2x$. With $\|\gamma\|_1 = 1$, $h(X) = 0$, and $\mathbf{E}\left[\log_2 \gamma(X)\right] = 1 - 1/(\ln 2)$, the resulting distortion is

$$D_{\text{vr}} \approx \frac{1}{12} \cdot 4e^{-2} \cdot 2^{-2R} \approx 0.045 \cdot 2^{-2R}.$$

Quantizers designed with the three derived optimal point densities are illustrated in Fig. 2 for rate $R = 4$. The functionally-optimized quantizers put more points at higher values of x, where the function varies more quickly. In addition, the variable-rate quantizer is allowed more points ($K = 21$) while meeting the rate constraint.

The interested reader can verify that D_{fr} and D_{vr} exactly match the performance obtained by designing optimal quantizers for $Y = X^2$. □

The example shows that even for univariate functions, there are benefits from functional quantization. While quantizing X instead of $g(X)$ seems naïve, as we move to the distributed multivariate case it will not be possible to compute the function before quantization. The approach of linearly approximating g will generalize to allow optimization of quantizers.

E. Discontinuous Functions

Our main result on univariate functional quantization, Theorem 4, assumes the continuity of g. One can effectively sidestep this assumption, but doing so requires the quantizer to be described more precisely than by a point density function alone.

For simplicity, assume f_X is strictly positive on $[0, 1]$. Suppose we were to allow g to have a point of discontinuity $x_0 \in (0, 1)$ with

$$c_0 = \lim_{\delta \to 0} |g(x_0 + \delta) - g(x_0 - \delta)| > 0.$$

The difficulty that arises is that if x_0 is an interior point of a partition cell S_i, this cell produces a component of the functional distortion proportional to $c^2 \mathbf{P}\left(X \in S_i\right)$. Since $c^2 \mathbf{P}\left(X \in S_i\right) = \Theta(K^{-1})$, it is not negligible in comparison to the (best case) $\Theta(K^{-2})$ functional distortion. Thus having a point of discontinuity of g in the interior of a partition cell disrupts the asymptotic distortion calculation (9).

The representation of quantizers by number of levels K and point density function λ does not allow us to prevent a point of discontinuity from falling in the interior of a partition cell. However, if we augment the description of the quantizer with specified partition boundaries, we can still obtain the distortion estimate (9).

Corollary 5: Suppose a K-level quantizer for a source $X \in [0,1]$ is described by point density function $\lambda(x)$. Further suppose that the source, quantizer, and function $g : [0,1] \to \mathbb{R}$ satisfy Assumptions HR1–3 and UF1–2 with the exception of discontinuities at M points $\{x_m\}_{m=1}^M$. Then a $(K+M)$-level quantizer obtained by adding partition cell boundaries at $\{x_m\}_{m=1}^M$ will have distortion

$$D = \mathbf{E}\left[(g(X) - g(\widehat{X}))^2\right] \approx \frac{1}{12K^2}\mathbf{E}\left[(\gamma(X)/\lambda(X))^2\right].$$

Proof: The proof is omitted, as it requires only minor modifications of the proof of Theorem 4 in Appendix A. ∎

In the sequel, we will not consider discontinuous functions. It seems that a multivariate extension of Corollary 5 would require points of discontinuity to be in the Cartesian product of finite sets of discontinuity for each variable. Such separable sets of points of discontinuity are not general and can be handled rather intuitively.

IV. Multivariate Functional Quantization

With Section III as a warm-up, we can now address the actual distributed functional scalar quantization problem. Let X_1^n be a random vector with joint pdf $f_{X_1^n}(x_1^n)$ defined over $[0,1]^n$, and let $g : [0,1]^n \to \mathbb{R}$ be the function of interest. As depicted in Fig. 1, each source X_j is quantized at rate R_j into $\widehat{X}_j = Q_j(X_j)$, separately, and an estimate $\widehat{g}(\widehat{X}_1^n)$ is formed at the decoder, where $\widehat{g}$ is the estimator function. We wish to design $\widehat{g}$ and $Q_j(X_j)$, $j = 1, 2, \ldots, n$, to minimize the functional distortion, $D = \mathbf{E}[(g(X_1^n) - \widehat{g}(\widehat{X}_1^n))^2]$.

A. Assumptions

As in Section III, we will impose restrictions on the function g and the joint distribution of X_1^n so that a local affine approximation is effective. For $j \in \{1, 2, \ldots, n\}$, let $\{\beta_i^{(j)}\}_{i \in \mathcal{I}^{(j)}}$ denote the quantization points and $\{S_i^{(j)}\}_{i \in \mathcal{I}^{(j)}}$ the partition cells of the quantizer Q_j. We require each partition to satisfy Assumption HR3. As a multivariate counterpart to Assumption HR1, we require:

HR1′. $f_{X_1^n}$ is smooth enough that it may be approximated as constant on each cell $S_{i_1}^{(1)} \times S_{i_2}^{(2)} \times \cdots \times S_{i_n}^{(n)}$ in the rectangular partition induced by all n quantizers together.[2]

To simplify the proof of the main result, we make a slightly stronger smoothness assumption on multivariate function g than in the previous section:

MF1. g is monotonic in each variable.

MF2. g is twice continuously differentiable on $[0,1]^n$.

With the monotonicity requirement, Lemma 3 applies separately to each quantizer to show that designing regular quantizers does not preclude optimality. We will analyze the case of $\widehat{g} = g$ and then formally justify this by showing that the difference in distortions between using $\widehat{g} = g$ and the optimal $\widehat{g}$ is asymptotically negligible (Theorem 7).

B. High-Resolution Distortion

Our main technical task in finding the optimal quantizers is to justify an approximation of the distortion in terms of point density functions. Since the quantization is distributed, our concept of functional sensitivity is now extended to each variable separately, with averaging over all the remaining variables.

Definition 2: The jth *functional sensitivity profile* of g is defined as

$$\gamma_j(x) = \left(\mathbf{E}\left[|g_j(X_1^n)|^2 \mid X_j = x\right]\right)^{1/2}$$

where $g_j(x_1^n)$ denotes $\partial g(x_1^n)/\partial x_j$.

Theorem 6: Suppose n sources $X_1^n \in [0,1]^n$ are quantized in a distributed manner with a K_j-level quantizer with point density λ_j applied to X_j. Further suppose that the source, quantizers, and function $g : [0,1]^n \to \mathbb{R}$ satisfy the assumptions of Section IV-A. Then

$$D = \mathbf{E}\left[(g(X_1^n) - g(\widehat{X}_1^n))^2\right] \approx \sum_{j=1}^n \frac{1}{12K_j^2}\mathbf{E}\left[\left(\frac{\gamma_j(X_j)}{\lambda_j(X_j)}\right)^2\right]. \tag{14}$$

Proof: See Appendix B. ∎

Theorem 7: Assume the conditions of Theorem 6, and denote the performance of the optimal estimator by

$$D_{\text{opt}} = \mathbf{E}\left[\left(g(X_1^n) - \mathbf{E}\left[g(X_1^n) \mid \widehat{X}_1^n\right]\right)^2\right].$$

Then $D \approx D_{\text{opt}}$.

Proof: See Appendix C. ∎

[2] See [27] for a discussion of this local uniformity condition.

C. Optimal Point Densities

Theorem 6 uses the functional sensitivity profiles to decouple our design problem into n separate problems of designing a single point density λ_j. Furthermore, each design problem (the minimization of a term of (14)) is of a familiar form. Thus we obtain the following theorem.

Theorem 8: For fixed $\{K_j\}_{j=1}^n$—corresponding to fixed-rate quantizers at specified rates—the distortion expression (14) is minimized by the choice

$$\lambda_j(x) = \frac{\left(\gamma_j^2(x) f_{X_j}(x)\right)^{1/3}}{\int \left(\gamma_j^2(t) f_{X_j}(t)\right)^{1/3}\, dt}, \quad j = 1,\, 2\, \ldots,\, n, \tag{15}$$

resulting in distortion

$$D \approx \sum_{j=1}^{n} \frac{1}{12K_j^2} \|\gamma_j^2 f_{X_j}\|_{1/3} = \frac{1}{12} \sum_{j=1}^{n} \|\gamma_j^2 f_{X_j}\|_{1/3} 2^{-2R_j}, \tag{16}$$

where $R_j = \log_2 K_j$ is the rate of Q_j. If $\sum_{j=1}^n R_j$ is fixed to nR with R large enough and no requirement that $\{2^{R_j}\}_{j=1}^n$ be integers, the minimum distortion

$$D \approx \frac{n}{12} \left(\prod_{j=1}^{n} \|\gamma_j^2 f_{X_j}\|_{1/3} \right)^{1/n} 2^{-2R} \tag{17}$$

is achieved with λ_js given by (15) and

$$R_j = R + \frac{1}{2} \log_2 \frac{\|\gamma_j^2 f_{X_j}\|_{1/3}}{\left(\prod_{k=1}^{n} \|\gamma_k^2 f_{X_k}\|_{1/3}\right)^{1/n}}, \quad j = 1,\, 2\, \ldots,\, n. \tag{18}$$

For fixed entropies $\{H(\widehat{X}_j)\}_{j=1}^n$ given by (4)—corresponding to variable-rate quantizers at specified resolutions—the distortion (14) is minimized by the choice

$$\lambda_j(x) = \frac{\gamma_j(x)}{\int \gamma_j(t)\, dt}, \quad j = 1,\, 2\, \ldots,\, n. \tag{19}$$

As long as each λ_j is positive wherever f_{X_j} is positive, the high-resolution rate analysis is valid, and the resulting distortion can be written as

$$D \approx \sum_{j=1}^{n} \frac{1}{12} \|\gamma_j\|_1^2\, 2^{2h(X_j) + 2\mathbf{E}[\log_2 \gamma_j(X_j)]}\, 2^{-2R_j}, \tag{20}$$

where $R_j = H(\widehat{X}_j)$ is the output entropy of Q_j. If $\sum_{j=1}^n R_j$ is fixed to nR and R is large enough, the minimum distortion

$$D \approx \frac{n}{12} \left(\prod_{j=1}^{n} \|\gamma_j\|_1^2\, 2^{2h(X_j) + 2\mathbf{E}[\log_2 \gamma_j(X_j)]} \right)^{1/n} 2^{-2R} \tag{21}$$

is achieved with λ_js given by (19) and

$$R_j = R + \frac{1}{2} \log_2 \frac{\|\gamma_j\|_1^2\, 2^{2h(X_j) + 2\mathbf{E}[\log_2 \gamma_j(X_j)]}}{\left(\prod_{k=1}^{n} \|\gamma_k\|_1^2\, 2^{2h(X_k) + 2\mathbf{E}[\log_2 \gamma_k(X_k)]}\right)^{1/n}}, \tag{22}$$

for $j = 1,\, 2\, \ldots,\, n$.

Proof: To prove (15), (16), (19), and (20), it suffices to note that minimizing the n terms of (14) separately gives problems identical to those in Section III.

Minimizing (16) through the choice of rates summing to nR is precisely addressed by Lemma 1; this yields (17)–(18). Bit allocation (22) and resulting distortion (21) similarly follow easily from (20). ■

D. Variable-Rate with Slepian–Wolf Coding

Distortion expressions (17) and (21) are minimum distortions subject to a sum-rate constraint. The individual rates given by $R_j = \log_2 K_j$ or by (4) implicitly specify no entropy coding or separate entropy coding of the $\widehat{X}_j$s, respectively.

If the $\widehat{X}_j$s are not independent—which is anticipated whenever the X_js are not independent—one may employ Slepian–Wolf coding of the $\widehat{X}_j$s without violating the distributed coding requirement implicit in Fig. 1. This lowers the total rate from $\sum_{j=1}^n H(\widehat{X}_j)$ to $H(\widehat{X}_1,\, \widehat{X}_2,\, \ldots,\, \widehat{X}_n)$. In this section we study how the inclusion of Slepian–Wolf coding affects the optimal quantizers and the resulting performance.

Following the development of (4) line-by-line gives a high-resolution joint entropy estimate

$$H(\widehat{X}_1^n) \approx h(X_1^n) + \sum_{j=1}^n \log_2 K_j + \sum_{j=1}^n \mathbf{E}\left[\log_2 \lambda_j(X_j)\right], \tag{23}$$

where $h(X_1^n)$ is the joint differential entropy of X_1^n. The distortion expression (14) does not depend on the presence or absence of Slepian–Wolf coding.

Theorem 9: The minimum of the distortion (14) over the choice of point densities $\{\lambda_j\}_{j=1}^n$ and resolutions $\{K_j\}_{j=1}^n$ subject to upper bound nR on joint entropy (23) is

$$D \approx \frac{n}{12}\left(2^{2h(X_1^n)} \prod_{j=1}^n \|\gamma_j\|_1^2\, 2^{2\mathbf{E}[\log_2 \gamma_j(X_j)]}\right)^{1/n} 2^{-2R}. \tag{24}$$

It may be attained by the point densities (19) and $\{K_j\}_{j=1}^n$ satisfying

$$\widetilde{R}_j = h(X_j \mid X_1^{j-1}) + \log_2 K_j + \mathbf{E}\left[\log_2 \lambda_j(X_j)\right], \tag{25}$$

where

$$\widetilde{R}_j = R + \frac{1}{2}\log_2 \frac{\|\gamma_j\|_1^2\, 2^{2\mathbf{E}[\log_2 \gamma_j(X_j)]}}{\left(\prod_{k=1}^n \|\gamma_k\|_1^2\, 2^{2\mathbf{E}[\log_2 \gamma_k(X_k)]}\right)^{1/n}}. \tag{26}$$

Proof: First suppose that the K_js are fixed. For clarity define

$$D_j = \frac{1}{12K_j^2}\mathbf{E}\left[\left(\frac{\gamma_j(X_j)}{\lambda_j(X_j)}\right)^2\right], \quad j = 1, 2\ \ldots, n.$$

Then since $nR = \sum_{j=1}^n \widetilde{R}_j$ and $D = \sum_{j=1}^n D_j$, we have n decoupled optimizations: for $j = 1, 2, \ldots, n$, minimize D_j subject to an upper bound on $\widetilde{R}_j$. The solution for any j is to use the point density (19), resulting in distortion component

$$D_j \approx \frac{1}{12}\|\gamma_j\|_1^2\, 2^{2h(X_j|X_1^{j-1}) + 2\mathbf{E}[\log_2 \gamma_j(X_j)]}\, 2^{-2R_j}.$$

Now minimizing $\sum_{j=1}^n D_j$ subject to a constraint on $\sum_{j=1}^n \widetilde{R}_j$ is addressed by Lemma 1. It yields (24) and (26) when one notes that the product $\prod_{j=1}^n 2^{2h(X_j|X_1^{j-1})}$ that appears in the product $\prod_{j=1}^n D_j$ is $2^{2h(X_1^N)}$, by the chain rule of differential entropy. ∎

Some remarks:

1) By comparing (24) to (21), we see that the inclusion of Slepian–Wolf coding has reduced the sum rate to achieve any given distortion by
$$\left(\sum_{j=1}^n h(X_j)\right) - h(X_1^n).$$
This is, of course, not unexpected as it represents the excess information in the product of marginal distributions as compared to the joint distribution. This has been termed the *multiinformation* [28] and equals the mutual information when $n = 2$.
2) The use of $h(X_j \mid X_1^{j-1})$ in (25) is somewhat arbitrary. It can be replaced by any achievable point on the Slepian-Wolf joint-entropy boundary. The optimal $\widetilde{R}_j$s and resulting distortion D would not be affected, but the K_js would change. One can interpret this as a flexibility in resolution allocation (slightly distinct from bit allocation) that can be used to control inaccuracies due to the high-resolution approximations.
3) The theorem seems to analytically separate correlations among sources from functional considerations, exploiting correlation even though the quantizers are regular. In reality, the binning introduced by Slepian–Wolf coding makes the quantizers effectively nonregular to remove redundancy between sources.

E. Relationship to Locally-Quadratic Distortion Measures

In [17], the authors consider the class of "locally-quadratic" distortion measures for variable-rate high-resolution quantization. They define locally-quadratic measures as those having the following two properties:

1) Let x be in $\mathbb{R}^n$. For y sufficiently close to x in the Euclidean metric, the distortion between x and y is well approximated by $\sum_{i=1}^n M_i(x)|x_i - y_i|^2$, where $M_i(x)$ is a positive scaling factor. In other words, the distortion is a space-varying non-isotropically scaled MSE.
2) The distortion between two points is zero if and only if the points are identical.

For these distortion measures, they consider high-resolution variable-rate regular quantization, generalize Bucklew's results [16] to non-functional distortion measures, and demonstrate the use of multidimensional companding functions to implement these quantizers. Of particular interest is the comparison they perform between joint vector quantization and separable scalar quantization. When Slepian–Wolf coding is employed for the latter, the scenario is similar to the developments of this section.

The source of this similarity is the implicit distortion measure we work with: $d_g(x,y) = |g(x) - g(y)|^2$. When x and y are very close to each other, Taylor approximation reduces this error to a quadratic form:

$$|g(x) - g(y)|^2 \approx \sum_{i=1}^{n} \left| \frac{\partial g(x_1^n)}{\partial x_i} \right|^2 |x_i - y_i|^2.$$

From this, one may obtain the same variable-rate Slepian–Wolf performance as (24) via the locally quadratic approach.

However, there are important differences between locally-quadratic distortion measures and the functional distortion measures we consider. First and foremost: a scalar function of n variables, $n > 1$, is *guaranteed* to have an uncountable number of pairs $x \neq y$ for which $g(x) = g(y)$ and therefore that $d_g(x,y) = 0$.[3] This violates the second condition of a locally-quadratic distortion measure, and the repercussions are felt most strikingly for non-monotonic functions—those for which regular quantizers are not necessarily optimal (see Section VI).

The second condition is also violated by functions that are not *strictly* monotonic in each variable; one finds that without strictness, variable-rate analysis of the centralized encoding problem is invalidated. Specifically, if the derivative vector

$$\left(\frac{\partial g(x_1^n)}{\partial x_1}, \frac{\partial g(x_1^n)}{\partial x_2}, \ldots, \frac{\partial g(x_1^n)}{\partial x_n} \right)$$

has nonzero probability of possessing a zero component, the expected variable-rate distortion as derived by both Bucklew and Linder *et al.* is $D = 0$, regardless of rate. This nonsensical answer arrives from the null derivative having violated the high-resolution approximation. Given that even the several example functions we consider in the following section fall into this trap, the raw centralized analysis has limited applicability to functional scenarios. In future work, generalizations of our results in Section VII may be able to address such deficiencies.

V. Examples

Before moving on to extensions of the basic theory, which complicate matters, we present a few examples to show how optimal ordinary scalar quantization and optimal DFSQ differ. We especially want to highlight a few simple examples in which performance scaling with respect to n differ greatly between ordinary and functionally-optimized quantization.

Example 2 (Linear function): Consider the function $g(x_1^n) = \sum_{j=1}^{n} \alpha_j x_j$ where the α_js are scalars. Then for any j, $\gamma_j(x) = |\alpha_j|$. Since $\gamma_j(x)$ does not depend on x, it has no influence on the optimal point density for either the fixed- or variable-rate case; see (15) and (19).

Although $\gamma_j(x)$ gives no information on which values of X_j are more important than others (or rather shows that they are all equally important) the set of γ_js shows the relative importance of the components. This is reflected in optimal bit allocations computed via (18) or (22). □

Example 3 (Maximum): Let the set of sources X_1^n be uniformly distributed on $[0,1]^n$ and hence mutually independent. Consider the function

$$g(x_1^n) = \max(x_1, x_2, \ldots, x_n).$$

Though very simple, this function is more interesting than a linear function because the derivative with respect to one variable depends sharply on all the others. The function is symmetric in its arguments, so for notational convenience consider only the design of the quantizer for X_1.

The partial derivative $g_1(x_1^n)$ is 1 where the maximum is x_1 and is 0 otherwise. Thus,

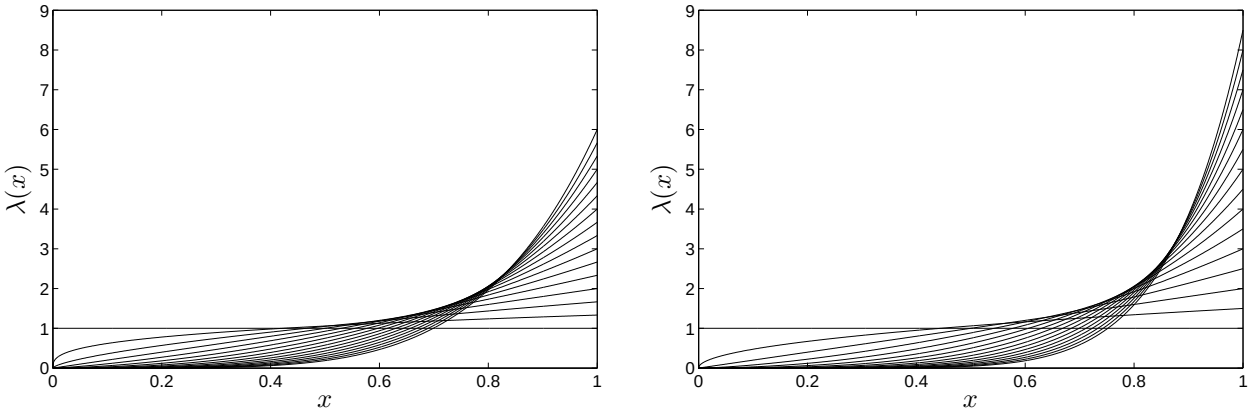
$$\begin{aligned} \gamma_1^2(x) &= \mathbf{E}\left[|g_1(X_1^n)|^2 \mid X_1 = x \right] \\ &= \mathbf{P}\left(\max(X_1^n) = X_1 \mid X_1 = x \right) \\ &= x^{n-1}, \end{aligned}$$

where the final step uses the probability of all $n-1$ variables X_2^n being less than x.

The optimal point density for fixed-rate quantization is found by evaluating (15) to be

$$\lambda_1(x) = \tfrac{1}{3}(n+2)x^{(n-1)/3}.$$

[3] If not, $\mathbb{R}^n$ has the same size as $\mathbb{Z} \times \mathbb{R}$, which is an absurdity.



(a) fixed-rate (b) variable-rate

Fig. 3. Optimal point densities for Example 3 (maximum), $n = 1, 2, \ldots, 16$. As n increases, the sensitivities $\gamma_j(x)$ become more unbalanced toward large x; this is reflected in the point densities, more so in the variable-rate case than in the fixed-rate case.

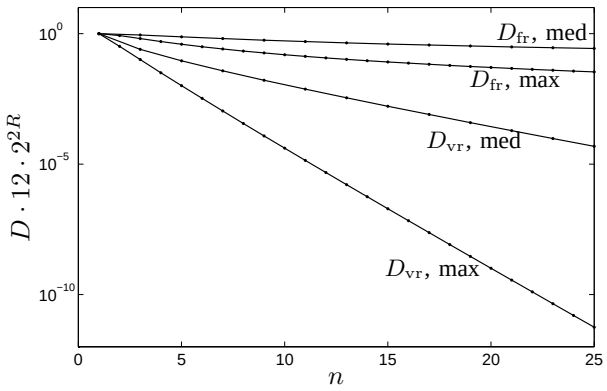


Fig. 4. Distortions of optimal fixed- and variable-rate functional quantizers for maximum and median functions from Examples 3 and 4. Shown is the dependence on the number of variables n; by plotting $D \cdot 12 \cdot 2^{2R}$ we see the performance relative to ordinary quantization.

The resulting distortion when each quantizer has rate R is found by evaluating (17) to be

$$\begin{aligned} D_{\mathrm{fr}} &\approx \frac{n}{12} \|\gamma_1^2\|_{1/3} \, 2^{-2R} = \frac{n}{12} \left(\frac{3}{n+2} \right)^3 2^{-2R} \\ &= \frac{9n}{4(n+2)^3} \, 2^{-2R}. \end{aligned}$$

The optimal point density for variable-rate quantization is found by evaluating (19) to be

$$\lambda_1(x) = \tfrac{1}{2}(n+1)x^{(n-1)/2}.$$

Substituting $\|\gamma_1\|_1 = 2/(n+1)$, $h(X_1) = 0$, and $2^{2\mathbf{E}[\log_2 \gamma_1(X_1)]} = e^{-n+1}$ into (21) gives

$$D_{\mathrm{vr}} \approx \frac{n}{12} \cdot \frac{4}{(n+1)^2} e^{-n+1} \, 2^{-2R} = \frac{en}{3(n+1)^2} e^{-n} \, 2^{-2R}.$$

The two computed distortions decrease sharply with n. This is in stark contrast to the results of ordinary quantization. When functional considerations are ignored, one optimally uses a uniform quantizer, resulting in $\mathbf{E}[(X_j - \widehat{X}_j)^2] \approx \frac{1}{12} 2^{-2R_j}$ for any component. Since the maximum is equal to one of the components, the functional distortion is $D_{\mathrm{ord}} \approx \frac{1}{12} 2^{-2R}$, unchanging with n.

The optimal point densities computed above are shown in Fig. 3. The distortions are presented along with the results of the following example in Fig. 4. □

Example 4 (Median): Let $n = 2m+1$, $m \in \mathbb{N}$, and again let the set of sources X_1^n be uniformly distributed on $[0,1]^n$. The function

$$g(x_1^n) = \mathrm{median}(x_1, x_2, \ldots, x_n)$$

provides a similar but more complicated example.

The partial derivative $g_1(x_1^n)$ is 1 where the median is x_1 and is 0 otherwise. Thus,

$$\begin{aligned}\gamma_1^2(x) &= \mathbf{E}\left[|g_1(X_1^n)|^2 \mid X_1 = x\right] \\ &= \mathbf{P}\left(\text{median}(X_1^n) = X_1 \mid X_1 = x\right) \\ &= \binom{2m}{m} x^m (1-x)^m,\end{aligned}$$

where the final step uses the binomial probability for the event of exactly m of the $2m$ variables X_2^n exceeding x.

The optimal point density for fixed-rate quantization is found by evaluating (15) to be

$$\lambda_1(x) = \frac{x^{m/3}(1-x)^{m/3}}{B(m/3+1, m/3+1)}$$

where B is the beta function. The resulting distortion when each quantizer has rate R is found by evaluating (17) to be

$$\begin{aligned}D_{\text{fr}} &\approx \frac{2m+1}{12} \|\gamma_1^2\|_{1/3} 2^{-2R} \\ &= \frac{2m+1}{12} \binom{2m}{m} \left(B\left(\frac{m}{3}+1, \frac{m}{3}+1\right)\right)^3 2^{-2R}.\end{aligned}$$

To understand the trend for large m, we can substitute in the Stirling approximations $\binom{2m}{m} \sim (m\pi)^{-1/2} 2^{2m}$ and

$$B(m/3+1, m/3+1) \sim \sqrt{6\pi/m}\, 2^{-(2m/3+3/2)}$$

to obtain

$$D_{\text{fr}} \sim \frac{m}{6} \frac{2^{2m}}{\sqrt{m\pi}} \left(\frac{6\pi}{m}\right)^{3/2} 2^{-(2m+9/2)} 2^{-2R} = \frac{\pi\sqrt{3}}{16m} 2^{-2R}.$$

The optimal point density for variable-rate quantization is found by evaluating (19) to be

$$\lambda_1(x) = \frac{x^m(1-x)^m}{B(m+1, m+1)}.$$

To evaluate the resulting distortion, note that

$$\|\gamma_1\|_1^2 = \binom{2m}{m} \left(B\left(\frac{m}{2}+1, \frac{m}{2}+1\right)\right)^2,$$

$h(X_1) = 0$, and $2^{2\mathbf{E}[\log_2 \gamma_1(X_1)]} = \binom{2m}{m} e^{-2m}$. Substituting into (21) gives

$$D_{\text{vr}} \approx \frac{2m+1}{12} \binom{2m}{m}^2 \left(B\left(\frac{m}{2}+1, \frac{m}{2}+1\right)\right)^2 e^{-2m} 2^{-2R}.$$

Using the approximation above for the binomial factor and

$$B(m/2+1, m/2+1) \sim \sqrt{4\pi/m}\, 2^{-(m+3/2)},$$

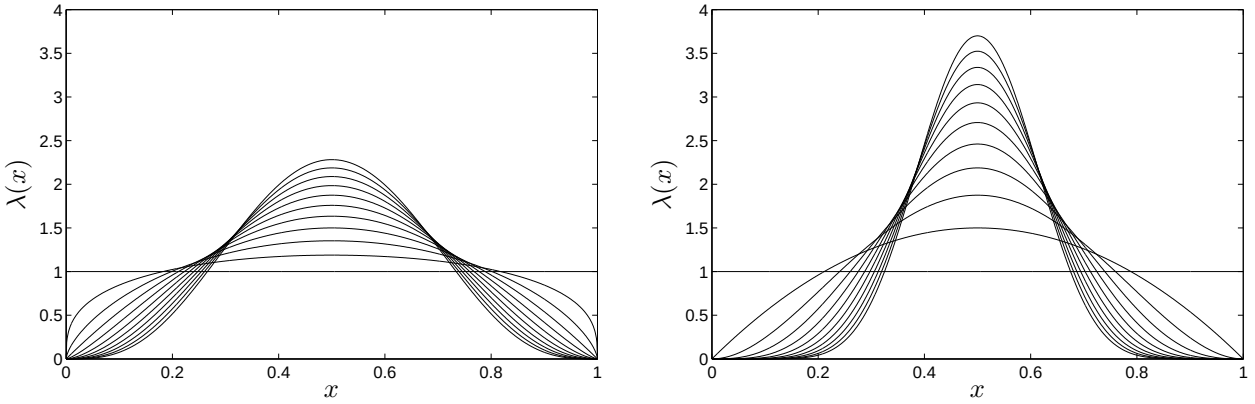
we obtain

$$\begin{aligned}D_{\text{vr}} &\sim \frac{m}{6} \frac{2^{4m}}{m\pi} \frac{4\pi}{m} 2^{-(2m+3)} e^{-2m} 2^{-2R} \\ &= \frac{1}{12m} \left(\frac{e}{2}\right)^{-2m} 2^{-2R}.\end{aligned}$$

The optimal point densities computed above are shown in Fig. 5. The distortions are presented along with the results of Example 3 in Fig. 4.

Note the following similarities to Example 3: D_{ord} is constant with respect to n, D_{fr} decays polynomially with n, and D_{vr} decays exponentially with n. □

Additional examples and details appear in [29].



(a) fixed-rate (b) variable-rate

Fig. 5. Optimal point densities for Example 4 (median), $n = 1, 3, \ldots, 21$. As n increases, the sensitivities $\gamma_j(x)$ become more unbalanced toward $x = 1/2$; this is reflected in the point densities, more so in the variable-rate case than in the fixed-rate case.

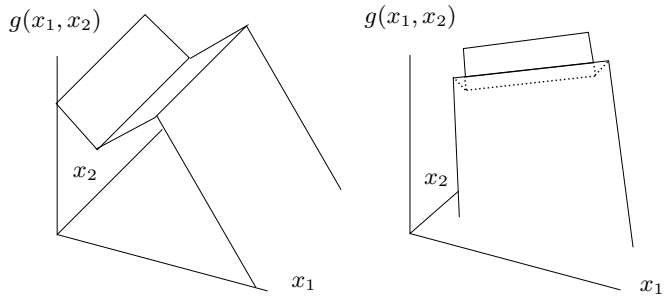


Fig. 6. Two versions of a function g of two variables are shown. The left g is separable and X_1 is best quantized by a non-regular quantizer; for the right function (a rotated version of the left), a regular quantizer is asymptotically optimal. This is due to the right function being "equivalence-free."

VI. Non-Monotonic Functions and Non-Regular Quantization

The high-resolution approach to quantizer optimization is inherently limited to the design of regular quantizers. In particular, a point density function describes only the quantizer point locations; the partition is implicit. The analysis of Section IV therefore gave us the best quantizers within the class of regular quantizers, and it was the restriction of attention to monotonic functions that ensured global optimality.

In this section we explore less restrictive alternatives to the monotonicity requirement. Specifically, we introduce the concept of *equivalence-free* and show that if a function has this property, then at a high enough rate, the optimal functional quantizers must be regular.

Fig. 6 illustrates the concept. The function on the left is aligned with the axes in the sense that $g(x_1, x_2)$ depends only on x_1. Since the dependence on x_1 is not monotonic, there are pairs $(x_1^\dagger, x_1^\ddagger)$ where $g(x_1^\dagger, x_2) = g(x_1^\ddagger, x_2)$ and thus the optimal quantizer at high enough resolution has $Q_1(x_1^\dagger) = Q_1(x_1^\ddagger)$, giving a non-regular quantizer. When the function is rotated as shown on the right, there continues to be a lack of monotonicity but at high enough rate it can no longer be exploited. For some fixed x_2 there may be pairs $(x_1^\dagger, x_1^\ddagger)$ such that $g(x_1^\dagger, x_2) = g(x_1^\ddagger, x_2)$, but since the equality does not hold for all x_2, at sufficiently high rate it will not pay to have $Q_1(x_1^\dagger) = Q_1(x_1^\ddagger)$.

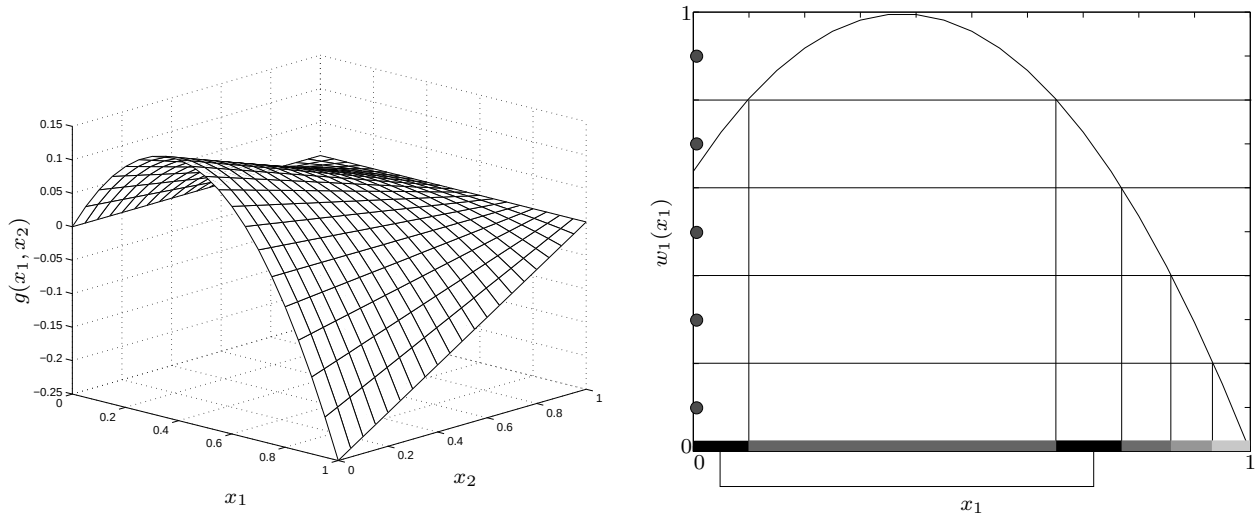
Our approach is to first create a model for high-resolution non-regular quantization, then to use this model to expand the class of functions for which regular quantization is optimal, and finally to construct asymptotically optimal non-regular quantizers when regularity is suboptimal.

A. High-Resolution Non-Regular Quantization

To accommodate non-regular quantization, we extend the compander-based model of quantization. Companding is to implement a non-uniform quantizer as $w^{-1}(q(w(x)))$ where w is a *compressor*, q is a uniform quantizer, and w^{-1} is an *expander*. The reader is referred to [2] for additional details and references to original sources.

In Bennett's development of optimal companding, it is natural to require w to be both monotonic and have a bounded derivative everywhere; the derivative $w'(x)$ is proportional to the quantizer point density $\lambda(x)$ that has been central in our development thus far. Whether we look at λ or w, the role is to set the relative sizes of the quantization cells.

Since optimal functional quantizers are not necessarily regular, we adapt the conventional development to implement non-regular quantizers.



(a) Function of interest (b) Generalized compressor

Fig. 7. Example of a generalized compressor $w_1(x_1)$ for a function $g(x_1, x_2)$ and the partition resulting from uniform quantization of $w_1(X_1)$. Notice that the compressor dictates both the relative sizes of cells and the binning of intervals of X values.

Definition 3: A function w is a *generalized compressor* if it is continuous, piecewise monotonic with a finite number of pieces, and has bounded derivative over each piece. The inverse map w^{-1} is called a *generalized expander* and is not necessarily a function.

As in ordinary companding, w and w^{-1} are used along with a uniform quantizer q as $w^{-1}(q(w(x)))$. The restriction to a finite number of pieces is a limitation on the types of non-regular quantizers that can be captured with this model: those for which every quantizer cell is a finite union of regular cells (intervals). Barring certain pathological situations, this restriction is reasonable.

Along with setting relative sizes of cells, w can now bin intervals together to provide for non-regularity. To illustrate this, let us briefly consider a simple example. Suppose that the pair (X_1, X_2) is uniformly distributed over $[0, 1]^2$, variable rate quantization is to be performed on both variables, and the function of interest is defined by

$$g(x_1, x_2) = x_1(\tfrac{3}{4} - x_1)(1 - x_2).$$

An optimal functional quantizer—a quantizer for X_1 to minimize $\mathbf{E}[(g(X_1, X_2) - g(\widehat{X}_1, \widehat{X}_2))^2]$—should bin together X_1 values that always yield the same $g(X_1, X_2)$. It can be seen from a plot of $g(x_1, x_2)$ (Fig. 7a) that the segment $X_1 \in [0, 3/8]$ is identical in this respect to the segment $X_1 \in [3/8, 3/4]$. This yields the constraint $w_1(x_1) = w_1(3/4 - x)$ for $x \in [0, 3/4]$. Furthermore, (19) sets the magnitude of the slope of w_1 in relation to the expected magnitude of the slope of g:

$$|w_1'(x)| \propto \frac{3}{4} - 2x_1.$$

This still leaves many choices for w_1, the most obvious being $w_1 = g(x_1, 0)$. A shifted and normalized version of this choice, along with the resulting quantizer, is drawn in Fig. 7b.

B. Equivalence-Free Functions

We now define a broad class of functions for which regular quantization is optimal at sufficiently high resolutions. Consider the distributed functional scalar quantization problem for a function $g(X_1^n)$ defined on $[0, 1]^n$ subject to mean-squared error distortion. We will focus on the design of the jth quantizer.

We require a set of definitions:

Definition 4: For any $s \neq t$ in the support of X_j, let

$$v_j(s, t) = \mathbf{E}\left[\mathrm{var}\left(g(X_1^n) \mid X_j \in \{s, t\}, \{X_i\}_{i \neq j}\right)\right].$$

If $v_j(s, t) = 0$ then (s, t) is a *functional equivalence in the jth variable*. If g has no functional equivalences in any of its variables, we say it is *equivalence-free*.

The theorem below demonstrates that for DFSQ with an equivalence-free function, quantizer regularity is asymptotically necessary for optimality. Specifically, non-regular quantization is shown to introduce a nonzero lower bound on the distortion, independent of rate. This is formalized with the aid of generalized companding.

Theorem 10: Let g be equivalence-free with respect to the distribution of X_1^n on $[0, 1]^n$. Suppose quantization of each X_j is performed as $\widehat{Y}_j = q(w_j(X_j))$ where w_j is a generalized compressor and q is a uniform quantizer. If there is an index j,

closed interval S, and function $t : \mathbb{R} \to \mathbb{R}$ such that $\mathbf{P}(X_j \in S) > 0$ and, for every $s \in S$, $s \neq t(s)$ and $w_j(s) = w_j(t(s))$, then the distortion has a positive, rate-independent lower bound.

Proof: See Appendix D. ■

The positive, rate-independent lower bound shows that the quantizer is suboptimal if the rate is sufficiently high; even naive uniform quantization will yield $D = O(2^{-2R})$ dependence on rate and thus will eventually outperform the non-regular quantizer. It should be noted, however, that the rate above which a non-regular quantizer is necessarily suboptimal has not been specified here.

When a function has equivalences, the best asymptotic quantization tactic is to design compressors that bin all the equivalent values in each variable but are otherwise monotonic.

VII. Don't-Care Intervals and Rate Amplification

Ordinary high-resolution analysis produces point-density functions that reflect the source distribution in the sense that optimal quantizers never have zero point density where there is nonzero probability density. In fact, having zero point density where there is nonzero probability density would contradict the conditions that validate the high-resolution analysis. The situation is more complicated in the functional setting since the optimal point densities depend on both the functional sensitivity profiles and the source distributions. As foreshadowed by the qualifications in Theorem 8, having zero functional sensitivity where the probability density is nonzero changes the optimal quantizers in the variable-rate case.

The following example illustrates the potential for failure of the analysis of Section IV-C. Note that the intricacies arise even with a univariate function.

Example 5: Let X have the uniform distribution over $[0, 1]$, and suppose the function of interest is $g(X) = \min(X, 1/2)$. It is clear that the optimal quantizer (for both fixed- and variable-rate) has uniform point density on $[0, 1/2]$. With the functional sensitivity profile given by

$$\gamma(x) = \begin{cases} 1, & \text{if } x < 1/2; \\ 0, & \text{otherwise,} \end{cases}$$

evaluating (10) and (12) is consistent with the intuitive result.

The distortion for the fixed-rate case obtained from (11) is $(1/12)(1/2)^3 2^{-2R}$. This is sensible since for half of the source values ($X > 1/2$) there is zero distortion by having a single codeword at $1/2$, whereas for the other half of the source values ($X < 1/2$), $2^R - 1$ codewords quantize a random variable uniformly distributed over $[0, 1/2]$.

The variable-rate case is problematic. Since $\mathbf{E}[\log_2 \gamma(X)] = -\infty$, evaluating (13) yields $D \approx 0$. The high-resolution *rate* analysis does not apply because the quantization is not fine over the full support of f_X. (The high-resolution *distortion* analysis is valid, as we will establish formally in Section VII-A.) The performance is easily described by considering the first representation bit to specify the event $A = \{X < 1/2\}$ or its complement. Since additional bits are useful only when A occurs, one can spend $2(R-1)$ bits in those cases to have an average expenditure of R bits. The resulting distortion is

$$\begin{aligned} D &= \mathbf{P}(A)\, D_{|A} + \mathbf{P}(A^c)\, D_{|A^c} \\ &\approx \tfrac{1}{2} \cdot \tfrac{1}{12} (\tfrac{1}{2})^2 2^{-2(2R-2)} + \tfrac{1}{2} \cdot 0 \;=\; \tfrac{1}{6} 2^{-4R}. \end{aligned}$$

Note that the exponent in the distortion–rate relationship has changed. □

In the example, there is an interval $X \in [1/2, 1]$ of source values that need not be distinguished for function evaluation. Let us define a term for such intervals before discussing the example further.

Definition 5: An interval $Z \subset [0, 1]$ is called a *don't-care interval* for the jth variable when the jth functional sensitivity γ_j is identically zero on Z, but the probability $\mathbf{P}(X_j \in Z)$ is positive.

In univariate FSQ, at high enough rates, each don't-care interval corresponding to a distinct value of the function should be allotted one codeword. This follows from reasoning similar to that given in Section VI-B and is illustrated by Example 5. In the fixed-rate case, the don't-care intervals simply occupy a few of the 2^R codewords and have a limited effect. Contrarily in the variable-rate case, the don't-care intervals produce a subset of source values that can be allotted very little rate. This gives more rate to be allotted outside the don't-care intervals and behavior we refer to as *rate amplification*. We demonstrate rate amplification for multivariate FSQ in Section VII-B after covering the distortion analysis and fixed-rate case in Section VII-A.

A. Distortion Analysis and Fixed-Rate Optimization

In the following analysis we will assume that the jth variable has a finite number M_j of don't-care intervals $\{Z_{j,1}, Z_{j,2}, \ldots, Z_{j,M_j}\}$. We also assume

$$\mathbf{P}(X_j \in Z_j) < 1 \quad \text{for } j = 1, 2 \ldots, n, \tag{27}$$

where $Z_j = \cup_{i=1}^{M_j} Z_{j,i}$ denotes the union of don't-care intervals for the jth variable. Without this, there is no improvement beyond M_j levels in representing X_j, so the high-resolution approach is wholly inappropriate. We will denote the event $X_j \notin Z_j$ by A_j.

For fixed-rate DFSQ, the optimal operational distortion–rate expression (16) remains valid when variable X_j has don't-care intervals, even though the optimal point density λ_j obtained from (15) is zero where f_{X_j} is nonzero (invalidating some arguments in Appendix B). Here we give an argument relying on an explicit characterization of the distortion similar to (14). At high enough rates, it is intuitive to allot a codeword of Q_j to each don't-care interval $Z_{j,i}$. The remaining $K_j - M_j$ codewords are assigned optimally to $[0,1] \setminus Z_j$ according to the basic theory developed in Section IV.

Theorem 11: Suppose n sources $X_1^n \in [0,1]^n$ are quantized in a distributed manner with a K_j-level quantizer with point density λ_j applied to X_j. Further suppose that the sources, quantizers, and function $g : [0,1]^n \to \mathbb{R}$ satisfy the assumptions of Section IV-A, with the exception that HR1′ need not hold for the jth variable where $\lambda_j = 0$. Finally, assume each source X_j has M_j don't-care intervals satisfying (27). Then the optimal point densities satisfy

$$\lambda_j(x) = 0 \quad \text{for all } x \in Z_j \tag{28}$$

and yield

$$\begin{aligned} D &= \mathbf{E}\left[(g(X_1^n) - g(\widehat{X}_1^n))^2\right] \\ &\approx \sum_{j=1}^{n} \frac{\mathbf{P}(A_j)}{12(K_j - M_j)^2} \mathbf{E}\left[\left(\frac{\gamma_j(X_j)}{\lambda_j(X_j)}\right)^2 \mid A_j\right]. \end{aligned} \tag{29}$$

The optimal point densities for fixed-rate quantization are given by (28) inside the don't-care intervals and by (15) outside of the don't-care intervals. These point densities yield

$$D \approx \sum_{j=1}^{n} \frac{1}{12(K_j - M_j)^2} \|\gamma_j^2 f_{X_j}\|_{1/3} \tag{30}$$

$$\approx \frac{1}{12} \sum_{j=1}^{n} \|\gamma_j^2 f_{X_j}\|_{1/3} 2^{-2R_j}. \tag{31}$$

Proof: See Appendix E. ∎

B. Variable-Rate and Rate Amplification

In the variable-rate case, it remains true that don't-care intervals should not be finely quantized (see (28)) and the distortion calculation (29) holds. The distinction from the fixed-rate case is that $X_j \in Z_j$ not only implies that the jth variable has limited impact on the distortion, but also that it can be allocated very little rate.

To formalize the analysis, we define discrete random variables to represent the events of source variables lying in don't-care intervals.

Definition 6: The random variable

$$I_j = \begin{cases} i, & \text{if } X_j \in Z_{j,i} \text{ for } i \in \{1, 2, \ldots, M_j\}; \\ 0, & \text{otherwise.} \end{cases}$$

is called the jth *don't-care variable*. The previously-defined event A_j can be expressed as $\{I_j = 0\}$.

At high enough rates, the jth encoder communicates I_j and in addition, *only when* $I_j = 0$, a fine quantization of X_j. The resulting performance is summarized by the following theorem.

Theorem 12: Under the conditions of Theorem 11, the optimal point densities for variable-rate quantization follow (19) and yield

$$\begin{aligned} D \approx \frac{1}{12} \sum_{j=1}^{n} \alpha_j^{-1} \|\gamma_j\|_1^2 \\ \times 2^{-2(\alpha_j(R - H(I_j)) + h(X_j|A_j) + \mathbf{E}[\log_2(\gamma_j(X_j))|A_j]}, \end{aligned} \tag{32}$$

where $\alpha_j = 1/\mathbf{P}(A_j)$ is the *amplification* of R_j.

Proof: See Appendix F. ∎

Some remarks:

1) The quantity $H(I_j)$ may be identified as the cost of communicating the indicator information to the decoder. The remaining rate, $R_j - H(I_j)$, is amplified by factor α_j because additional description of X_j is useful only when $X_j \notin Z_j$. The amplification shows that the standard -6 dB/bit distortion decay may be exceeded in the presence of don't-care regions.
2) At moderate rates, it may not be optimal to communicate I_j losslessly, and it may be beneficial to include X_j values with small but positive γ_j in don't-care intervals. Study of this topic is left for specific applications.

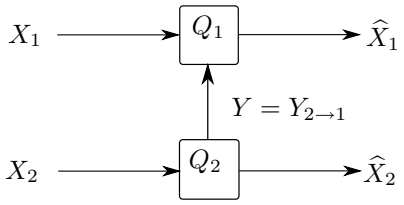


Fig. 8. Suppose the encoder for X_2 could send a bit to the encoder for X_1. Is there any benefit? How does it compare to sending an additional bit to the decoder?

3) The rate amplification we have seen in the variable-rate case and the relative lack of importance of don't-care intervals in the fixed-rate case have a close analogy in ordinary lossy source coding. Suppose a source X is a mixed random variable with an M-value discrete component and a continuous component. High-resolution quantization of X will allocate one level to each discrete value and the remaining levels to the continuous component. The discrete component changes the constant factor in $\Theta(2^{-2R})$ fixed-rate operational distortion–rate performance while it changes the decay rate in the variable-rate case. See [30] for related rate–distortion (rather than high-resolution quantization) results.

VIII. Chatting Encoders

Our final variation on the basic theory of distributed functional scalar quantization is to allow limited communication between the encoders. How much can the distortion be reduced via this communication? Echoing the results of the previous section, we will find dramatically different answers in the fixed- and variable-rate cases.

For notational convenience, we will fix the communication to be from encoder 2 to encoder 1 though the number of source variables n remains general. In accordance with the block diagram of Fig. 8, the information $Y = Y_{2\to 1}$ must be conditionally independent of X_1 given X_2. We first consider the case where Y is a single bit.

In this section, we express the functional distortion as

$$D \approx \frac{1}{12}\sum_{j=1}^{n} D_j 2^{-2R_j},$$

where various expressions for D_j have been found for different scenarios, including (16), (20), (31), and (32). At issue is how D_1 is affected by Y; the other D_js are obviously not affected.

A. Fixed-Rate Quantization

In general, the availability of a single bit Y causes one to choose between two potentially-different quantizers $Q_{1|Y=0}$ $Q_{1|Y=1}$ in the quantization of X_1. We express the optimal quantizers and the resulting distortion contribution D_1 by way of the following concept.

Definition 7: The jth *conditional functional sensitivity profile* of g given $Y = y$ is defined as

$$\gamma_{j|Y}(x \mid y) = \left(\mathbf{E}\left[|g_j(X_1^n)|^2 \mid X_j = x, Y = y\right]\right)^{1/2}$$

where $g_j(x_1^n)$ denotes $\partial g(x_1^n)/\partial x_j$.

Now several results follow by analogy with Theorem 8. For the case of $Y = y$, the optimal point density is given by

$$\lambda_{1|Y}(x \mid y) = \frac{\left(\gamma_{1|Y}^2(x \mid y) f_{X_1|Y}(x \mid y)\right)^{1/3}}{\int \left(\gamma_{1|Y}^2(t \mid y) f_{X_1|Y}(t \mid y)\right)^{1/3} dt}$$

resulting in conditional distortion contribution

$$\frac{1}{12K_1^2}\left\|\gamma_{1|Y=y}^2 f_{X_1|Y=y}\right\|_{1/3}.$$

Combining the two possibilities for Y via total expectation gives

$$D_1 = \sum_{y=0}^{1} \mathbf{P}(Y = y)\left\|\gamma_{1|Y=y}^2 f_{X_1|Y=y}\right\|_{1/3}. \tag{33}$$

From this expression we reach an important conclusion on the affect of the chatting bit Y.

Theorem 13: For fixed-rate quantization, communication of one bit of information from decoder 2 to decoder 1 can at most reduce D_1 by a factor of 4.

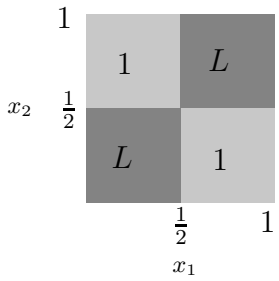


Fig. 9. Illustration for Example 6. Shown is the unit square $[0,1]^2$ with quadrants marked with the value of $g_1(x_1, x_2)$.

Proof: From Theorem 8, the distortion contribution analogous to (33) without the chatting bit Y is $\left\|\gamma_1^2 f_{X_1}\right\|_{1/3}$. Thus the fact we wish to prove is a statement about $\mathcal{L}^{1/3}$ pseudonorms of surrogate densities and their conditional forms.

We proceed as follows:

$$\begin{aligned}
D_1 &= \sum_{y=0}^{1} \left\| \mathbf{P}\left(Y=y\right) \gamma_{1|Y}^2(x \mid y) f_{X_1|Y}(x \mid y) \right\|_{1/3} \\
&\overset{(a)}{\geq} \frac{1}{4} \left\| \sum_{y=0}^{1} \mathbf{P}\left(Y=y\right) \gamma_{1|Y}^2(x \mid y) f_{X_1|Y}(x \mid y) \right\|_{1/3} \\
&= \frac{1}{4} \left\| f_{X_1}(x) \sum_{y=0}^{1} \frac{\mathbf{P}\left(Y=y\right) f_{X_1|Y}(x \mid y)}{f_{X_1}(x)} \gamma_{1|Y}^2(x \mid y) \right\|_{1/3} \\
&\overset{(b)}{=} \frac{1}{4} \left\| f_{X_1}(x) \sum_{y=0}^{1} \mathbf{P}\left(Y=y | X_1=x\right) \gamma_{1|Y}^2(x \mid y) \right\|_{1/3} \\
&\overset{(c)}{=} \frac{1}{4} \left\| f_{X_1}(x) \gamma_1^2(x) \right\|_{1/3}.
\end{aligned}$$

Step (a) uses a quasi-triangle inequality stated and proved in Appendix G; (b) is an application of Bayes's Rule; and (c) is based on an evaluation of the (unconditional) functional sensitivity via the total expectation theorem with conditioning on Y. This proves the theorem. ∎

Note that the result of Theorem 13 may be iterated for multiple bits of side information Y and that a factor of 4 reduction in D_1 per bit of communication may be *guaranteed* if the bits are instead put towards communication from encoder 1 to the decoder. These observations yield the following corollary.

Corollary 14: For fixed-rate functional quantization, communication of R additional bits between encoders performs *at best* as well as communication of R additional bits to the centralized decoder.

In general, the idea that bits from encoder 2 to encoder 1 are as good as bits from encoder 1 to the decoder is optimistic. In particular, if $\mathbf{E}\left[\gamma_1^2(X_1)\right] > 0$, then D_1 is bounded away from zero for any amount of communication from encoder 2 to encoder 1.

B. Variable-Rate Quantization

In a variable-rate scenario, the rate R_1 could be made to depend on the chatting bit Y, introducing a bit allocation problem between the cases of $Y=0$ and $Y=1$. Even without such dependence, we can demonstrate that the bit Y can reduce the first variable's contribution to the functional distortion by an arbitrary factor.

Analogous to (33),

$$D_1 = \sum_{y=0}^{1} \mathbf{P}\left(Y=y\right) \left\|\gamma_{1|Y=y}\right\|_1^2 2^{2h(X_1|Y=y)+2\mathbf{E}\left[\log_2 \gamma_{1|Y=y}(X_1)\right]} \tag{34}$$

by comparison with (20). In contrast to the $\mathcal{L}^{1/3}$ pseudonorms in (33), this linear combination can be arbitrarily smaller than

$$\|\gamma_1\|_1^2 \, 2^{2h(X_1)+2\mathbf{E}[\log_2 \gamma_1(X_1)]}.$$

We demonstrate this through a simple example.

Example 6: Let sources X_1 and X_2 be uniformly distributed on $[0,1]^2$. We specify the function of interest g through its partial derivatives. Let $g_2(x_1, x_2) = 1$ for all (x_1, x_2) and let $g_1(x_1, x_2)$ be piecewise constant as shown in Fig. 9, where L is a positive constant.

We can easily derive the first functional sensitivity profile of g to be

$$\gamma_1(x) = \sqrt{\tfrac{1}{2}(L^2+1)}.$$

This also allows us to find the distortion contribution factor D_1 without chatting to be

$$D_1 = \tfrac{1}{4}(L^2+1)^2.$$

In this example, one bit about X_2 is enough to allow the encoder for X_1 to perfectly tailor its point density to match the sensitivity of g at (X_1, X_2). Of course, the chatting bit should simply be

$$Y = \begin{cases} 0, & \text{if } X_2 > 1/2; \\ 1, & \text{otherwise.} \end{cases}$$

The first conditional functional sensitivity profiles for g are then

$$\gamma_{1|Y}(x \mid y) = \begin{cases} 1, & \text{for } Y = 0 \text{ and } X_1 \le 1/2 \\ & \text{or } Y = 1 \text{ and } X_1 > 1/2; \\ L, & \text{otherwise.} \end{cases}$$

Now for either value of y, we have $\int_0^1 \gamma_{1|Y}(x \mid y)\, dx = \frac{1}{2}(L+1)$ and $\mathbf{E}\left[\log_2 \gamma_{1|Y=y}(X_1)\right] = \frac{1}{2}\log_2 L$. Thus, evaluating (34) gives

$$D_1 = \tfrac{1}{4}(L+1)^2 L.$$

This is smaller than the D_1 with no chatting by about a factor of L. The performance gap can be made arbitrarily large by increasing L—all from a single bit of information communicated between encoders. □

C. Comparison with Non-Functional Source Coding

The results of this section are strikingly different from those of ordinary source coding. Consider first the discrete scenario in which we with to recreate X_1^n perfectly at the decoder. Can communication between encoders enable a reduction in the rate of communication to the decoder? According to Slepian and Wolf, the answer is a resounding "no." Even in the case of unlimited collaboration via fused encoders, the minimum sum rate to the decoder remains unchanged.

How about in lossy source coding? If quantization is variable-rate and Slepian–Wolf coding is employed on the quantization indices, no gains are possible from encoder interactions. This is a consequence of the work of Rebollo-Monedero *et al.* [9] on high-resolution Wyner–Ziv coding, where it is shown that there is no gain from supplying the source encoder with the decoder side information.

IX. Summary

We have developed asymptotically-optimal designs of functional quantizers using high-resolution quantization theory. This has shown that accounting for a function while quantizing a source can lead to arbitrarily large improvements in distortion. In certain scenarios (Section V), this improvement can grow exponentially with the number of sources. In others (Section VII), it can grow exponentially with rate.

Additionally, our study of functional quantization has highlighted some striking distinctions between fixed- and variable-rate cases:

1) For certain simple functions of order statistics, distortion relative to ordinary quantization falls polynomially with the number of sources in the fixed-rate case, whereas in the variable-rate case it falls exponentially.
2) The distortion associated with fixed-rate quantizers will always exhibit -6 dB/bit rate dependence at high rates, whereas the decay of distortion can be faster in some variable-rate cases.
3) Information sent from encoder-to-encoder can lead to arbitrarily-large improvements in distortion for variable-rate, whereas for fixed-rate this information can be no more useful than if it were sent to the decoder.

The second and third of these have extensions or analogues beyond functional quantization. Rate amplification is a feature of quantizing sources with mixed distributions, and the results on chatting encoders continue to hold when the function g is the identity operation.

APPENDIX A
PROOF OF THEOREM 4

The distortion can be written as

$$\begin{aligned} D &= \mathbf{E}\left[(g(X)-g(\widehat{X}))^2\right] \\ &= \sum_{i\in\mathcal{I}} \mathbf{E}\left[(g(X)-g(\beta_i))^2 \mid X\in S_i\right]\mathbf{P}\left(X\in S_i\right) \end{aligned} \tag{35}$$

by the law of total expectation. The desired expression (9) will be obtained by approximating the well-behaved terms in (35) and showing that the remaining terms can be safely ignored. Let $\mathcal{B}\subset\mathcal{I}$ be comprised of the indices i for which $g'(x)$ and $g''(x)$ are bounded for all $x\in S_i$.

Well-behaved terms: Let $i\in\mathcal{B}$. Then for $x\in S_i$,

$$g(x) = g(\beta_i) + g'(\beta_i)(x-\beta_i) + R_i(x)$$

where $R_i(x)$ is a remainder function bounded pointwise by (8). Now expand the conditional expectation in (35) as

$$\begin{aligned} &\mathbf{E}\left[(g(X)-g(\beta_i))^2 \mid X\in S_i\right] \\ &= \mathbf{E}\left[(g'(\beta_i)(X-\beta_i))^2 \mid X\in S_i\right] \\ &\quad + \mathbf{E}\left[2g'(\beta_i)(X-\beta_i)R_i(X) + R_i^2(X) \mid X\in S_i\right] \end{aligned}$$

The first term is easily approximated by high-resolution analysis and we wish to show that the second is asymptotically negligible.

For the first term, note that the approximate linearity of g on S_i implies we should place β_i at the center of S_i. Furthermore, the length of S_i is approximately $(K\lambda(\beta_i))^{-1}$ and X is conditionally approximately uniform on S_i. Thus the first term is $\frac{1}{12}(g'(\beta_i))^2(K\lambda(\beta_i))^{-2}$.

To bound the second term, note that

$$|2g'(\beta_i)(x-\beta_i)R_i(x) + R_i^2(x)|$$

has a uniform $O(\mathrm{length}(S_i)^3)$ bound for $x\in S_i$; this follows from the bounded derivatives in Assumption UF2 and bound (8). This makes the second term negligible in comparison to the first term, which is $\Theta(\mathrm{length}(S_i)^2)$.

Other terms: We now wish to show that the $i\in\mathcal{I}\setminus\mathcal{B}$ terms in (35) can be safely ignored. We do not have differentiability at $x\in S_i$, but the continuity of g prevents anything too bad from happening. Continuity on a closed interval implies uniform continuity, so there exists a finite constant c such that $|g(x)-g(\beta_i)| < c|x-\beta_i|$ for $x\in S_i$. The conditional expectation is thus bounded by $c^2(\mathrm{length}(S_i))^2$, and

$$\begin{aligned} &\sum_{i\in\mathcal{I}\setminus\mathcal{B}} \mathbf{E}\left[(g(X)-g(\beta_i))^2 \mid X\in S_i\right]\mathbf{P}\left(X\in S_i\right) \\ &\quad\leq\quad |\mathcal{I}\setminus\mathcal{B}|\cdot c^2\cdot \max_i(\mathrm{length}(S_i))^2 \end{aligned} \tag{36}$$

by replacing each term with an upper bound that does not even account for $\mathbf{P}\left(X\in S_i\right)\ll 1$. Thus at high resolution these terms may be ignored.

Final expression: We are now left with

$$\begin{aligned} D &\approx \sum_{i\in\mathcal{B}} \frac{1}{12}(g'(\beta_i))^2(K\lambda(\beta_i))^{-2}\mathbf{P}\left(X\in S_i\right) \\ &= \frac{1}{12K^2}\sum_{i\in\mathcal{B}}\left(g'(\beta_i)/\lambda(\beta_i)\right)^2\mathbf{P}\left(X\in S_i\right) \\ &\overset{(a)}{\approx} \frac{1}{12K^2}\int_{x\in S_i, i\in\mathcal{B}}\left(\gamma(x)/\lambda(x)\right)^2 f_X(x)\,dx \\ &\overset{(b)}{\approx} \frac{1}{12K^2}\mathbf{E}\left[\left(\gamma(X)/\lambda(X)\right)^2\right], \end{aligned}$$

where (a) is a standard high-resolution approximation; and (b) follows from $\mathbf{P}\left(X\in\cup_{i\in\mathcal{I}\setminus\mathcal{B}}S_i\right)\to 0$ and (36).

APPENDIX B
PROOF OF THEOREM 6

We wish to develop the estimate (14), which relates functional distortion to functional sensitivity profiles. The introduction of functional sensitivity profiles is motivated by the Taylor series approximation of g. Our main task is thus to show that the error in Taylor series approximation becomes negligible under the high-resolution assumptions of Section IV-A.

Recall the notation from Section IV-A: For $j \in \{1, 2, \ldots, n\}$, the quantization points of quantizer Q_j are denoted $\{\beta_i^{(j)}\}_{i \in \mathcal{I}^{(j)}}$ and the partition cells are denoted $\{S_i^{(j)}\}_{i \in \mathcal{I}^{(j)}}$. To simplify expressions below, let

$$\mathcal{I}_1^n = \mathcal{I}^{(1)} \times \mathcal{I}^{(2)} \times \cdots \times \mathcal{I}^{(n)},$$

$$\beta_{i_1^n} = \left(\beta_{i_1}^{(1)}, \beta_{i_2}^{(2)}, \ldots, \beta_{i_n}^{(n)}\right),$$

and

$$S_{i_1^n} = S_{i_1}^{(1)} \times S_{i_2}^{(2)} \times \cdots \times S_{i_n}^{(n)}.$$

Then we can express the distortion as

$$\begin{aligned} D &= \mathbf{E}\left[(g(X_1^n) - g(\widehat{X}_1^n))^2\right] \\ &= \sum_{i_1^n \in \mathcal{I}_1^n} \mathbf{E}\left[(g(X_1^n) - g(\beta_{i_1^n}))^2 \mid X_1^n \in S_{i_1^n}\right] \mathbf{P}\left(X_1^n \in S_{i_1^n}\right) \end{aligned} \tag{37}$$

by the law of total expectation. We wish to approximate the conditional expectations in this sum based on linear approximation of g within cell $S_{i_1^n}$, and we require vanishing relative error as resolution increases.

By Taylor's theorem,

$$g(x_1^n) = g(\beta_{i_1^n}) + \sum_{j=1}^{n} \left(x_j - \beta_{i_j}^{(j)}\right) g_j(\beta_{i_1^n}) + R_{i_1^n}(x_1^n) \tag{38}$$

where g_j denotes $\partial g / \partial x_j$ and the remainder term $R_{i_1^n}(x_1^n)$ is small near $\beta_{i_1^n}$. Specifically,

$$\left|R_{i_1^n}(x_1^n)\right| \le \frac{1}{2} \sum_{j=1}^{n} \sum_{k=1}^{n} (x_j - \beta_{i_j}^{(j)})(x_k - \beta_{i_k}^{(k)}) \max |g_{jk}(\xi_1^n)|$$

where $g_{jk}(x_1^n)$ denotes $\partial^2 g(x_1^n) / \partial x_j \partial x_k$ and the maximum is over ξ_1^n on the line connecting x_1^n and $\beta_{i_1^n}$. Assumption MF2 (that g is twice continuously differentiable) and the compactness of $[0,1]^n$ implies that there is an upper bound

$$|g_{jk}(x_1^n)| \le c \qquad \text{for all } x_1^n \in [0,1]^n.$$

Thus, for $x_1^n \in S_{i_1^n}$,

$$\left|R_{i_1^n}(x_1^n)\right| \le \tfrac{1}{2} c n^2 \max_{i,j} \left(\text{length}^2\left(S_i^{(j)}\right)\right). \tag{39}$$

Using the Taylor expansion with remainder,

$$\begin{aligned} &\mathbf{E}\left[(g(X_1^n) - g(\beta_{i_1^n}))^2 \mid X_1^n \in S_{i_1^n}\right] = \\ &\quad \mathbf{E}\left[\left(\sum_{j=1}^{n} \left(X_j - \beta_{i_j}^{(j)}\right) g_j(\beta_{i_1^n})\right)^2 \mid X_1^n \in S_{i_1^n}\right] + \\ &\quad \mathbf{E}\left[2\left(\sum_{j=1}^{n} \left(X_j - \beta_{i_j}^{(j)}\right) g_j(\beta_{i_1^n})\right) R_{i_1^n}(X_1^n) \mid X_1^n \in S_{i_1^n}\right] + \\ &\quad \mathbf{E}\left[R_{i_1^n}^2(X_1^n) \mid X_1^n \in S_{i_1^n}\right] \end{aligned} \tag{40}$$

Paralleling the development in Appendix A, the first term yields the desired approximation and the second and third terms are asymptotically negligible.

The first term can be evaluated under the assumptions of Section IV-A. Since $f_{X_1^n}$ is approximately constant on $S_{i_1^n}$ (Assumption HR1′) and the function is approximately affine on $S_{i_1^n}$, we can take $\beta_{i_1^n}$ to be the center of $S_{i_1^n}$. Then, conditioned

on $X_1^n \in S_{i_1^n}$, we have that the random variables $\{X_j - \beta_{i_j}^{(j)}\}_{j=1}^n$ are mutually uncorrelated. So all cross terms in the conditional expectation are zero, and

$$\begin{aligned}
&\mathbf{E}\left[\left(\sum_{j=1}^n \left(X_j - \beta_{i_j}^{(j)}\right) g_j(\beta_{i_1^n})\right)^2 \mid X_1^n \in S_{i_1^n}\right] \\
&\quad\approx \sum_{j=1}^n g_j^2(\beta_{i_1^n})\mathbf{E}\left[\left(X_j - \beta_{i_j}^{(j)}\right)^2 \mid X_1^n \in S_{i_1^n}\right] \qquad (41)\\
&\quad\approx \frac{1}{12}\sum_{j=1}^n g_j^2(\beta_{i_1^n})\left(K_j\lambda_j(\beta_{i_j}^{(j)})\right)^{-2}, \qquad (42)
\end{aligned}$$

where the last step uses the length of $S_{i_j}^{(j)}$ in the usual way.

Bounding the second and third terms of (40) is easy because we have already shown that $|R_{i_1^n}(x_1^n)| = O(\ell^2)$ on $S_{i_1^n}$, where ℓ denotes the maximum of the lengths of the sides of $S_{i_1^n}$. We thus have that the second and third term of (40) are together $O(\ell^3)$, which is negligible since the distortions we obtain are $O(\ell^2)$.

Having simplified (40) to (42), we substitute in (37) to make the final computations:

$$\begin{aligned}
D &\approx \sum_{i_1^n \in \mathcal{I}_1^n} \frac{1}{12}\sum_{j=1}^n g_j^2(\beta_{i_1^n})\left(K_j\lambda_j(\beta_{i_j}^{(j)})\right)^{-2}\mathbf{P}\left(X_1^n \in S_{i_1^n}\right)\\
&= \sum_{j=1}^n \frac{1}{12K_j^2}\sum_{i_1^n \in \mathcal{I}_1^n}\left(g_j(\beta_{i_1^n})/\lambda_j(\beta_{i_j}^{(j)})\right)^2\mathbf{P}\left(X_1^n \in S_{i_1^n}\right)\\
&\overset{(a)}{\approx} \sum_{j=1}^n \frac{1}{12K_j^2}\sum_{i_1^n \in \mathcal{I}_1^n}\left(g_j(X_1^n)/\lambda_j(X_j)\right)^2\mathbf{P}\left(X_1^n \in S_{i_1^n}\right)\\
&\overset{(b)}{\approx} \sum_{j=1}^n \frac{1}{12K_j^2}\int_{[0,1]^n}\left(g_j(x_1^n)/\lambda_j(x_j)\right)^2 f_{X_1^n}(x_1^n)\,dx_1^n\\
&\overset{(c)}{\approx} \sum_{j=1}^n \frac{1}{12K_j^2}\mathbf{E}\left[\left(\frac{\gamma_j(X_j)}{\lambda_j(X_j)}\right)^2\right],
\end{aligned}$$

where (a) uses that at high resolution, g_j and λ_j are approximately constant in a partition cell; (b) is the standard association of a sum with an integral; and (c) follows from first integrating over the $n-1$ variables excluding j to get squared functional sensitivity profiles in the integrand and then integrating over variable j.

Appendix C
Proof of Theorem 7

To minimize functional MSE, the optimal estimator clearly should compute the conditional expectation of $g(X_1^n)$ given the received codewords:

$$\widehat{g}(\beta_{i_1^n}) = \mathbf{E}\left[g(X_1^n) \mid X_1^n \in S_{i_1^n}\right],$$

where we have used notation from Appendix B. In analogy to (37),

$$\begin{aligned}
&D_{\text{opt}} = \qquad (43)\\
&\sum_{i_1^n \in \mathcal{I}_1^n}\mathbf{E}\left[(g(X_1^n) - \widehat{g}(\beta_{i_1^n}))^2 \mid X_1^n \in S_{i_1^n}\right]\mathbf{P}\left(X_1^n \in S_{i_1^n}\right).
\end{aligned}$$

Our goal is to show that $D - D_{\text{opt}}$ is asymptotically negligible, and we will do this by subtracting (43) from (37) and bounding each term. For this, we would like the conditional expectation of

$$\begin{aligned}
&(g(X_1^n) - g(\beta_{i_1^n}))^2 - (g(X_1^n) - \widehat{g}(\beta_{i_1^n}))^2\\
&\quad= \underbrace{(g(\beta_{i_1^n}) - \widehat{g}(\beta_{i_1^n}))}_{A}\underbrace{((g(\beta_{i_1^n}) - g(X_1^n))}_{B}\\
&\qquad+\underbrace{(\widehat{g}(\beta_{i_1^n}) - g(X_1^n)))}_{C}
\end{aligned}$$

to be small. We would like to obtain an $o(\ell^2)$ bound, where ℓ is the maximum of the lengths of the sides of $S_{i_1^n}$ as in Appendix B; since the distortion is $\Theta(\ell^2)$, this will make the suboptimality of the estimator g negligible.

To bound A, we first determine how close $\widehat{g}(\beta_{i_1^n})$ is to $g(\beta_{i_1^n})$, using the Taylor expansion with remainder derived in Appendix B. Computing the conditional expectation of (38) gives

$$\widehat{g}(\beta_{i_1^n}) = g(\beta_{i_1^n}) + \int_{S_{i_1^n}} R_{i_1^n}(x_1^n) f_{X_1^n|\{X_1^n \in S_{i_1{}^n}\}}(x_1^n) dx_1^n$$

where the first-order terms integrate to zero because of assumption HR1′. Now we can use (39) to conclude $|A| \leq \frac{1}{2} c n^2 \ell^2$.

To bound B and C, note that the conditions on g imply that it is Lipschitz continuous under any metric on the domain. Denoting the Lipschitz constant by L and using the ∞-norm for convenience, we obtain $|B| \leq L\ell$. Also, using the intermediate value theorem to argue that $\widehat{g}(\beta_{i_1^n}) = g(\xi_1^n)$ for some $\xi_1^n \in S_{i_1^n}$, we obtain $|C| \leq L\ell$.

Putting our calculations together, every conditional expectation that appears in the term-by-term difference between (43) and (37) is bounded by $(\frac{1}{2} c n^2 \ell^2)(2L\ell)$. Thus

$$D - D_{\text{opt}} \leq c n^2 L \ell^3,$$

which is asymptotically negligible.

Appendix D
Proof of Theorem 10

The theorem asserts that when the function is equivalence-free, w_j failing to be one-to-one on the support of X_j creates a component of the distortion that cannot be eliminated by quantizing more finely. The proof here lower-bounds the distortion by focusing on the contribution from just the jth variable. The bound is especially crude because it is based on observing $\{X_i\}_{i \neq j}$ and $w_j(X_j)$ without quantization and it uses only the contribution from $X_j \in S \cup t(S)$.

We wish to first bound the functional distortion in terms of a contribution from the jth variable:

$$\begin{aligned}
D &\overset{(a)}{\geq} \mathbf{E}\left[\operatorname{var}(g(X_1^n) \mid \widehat{Y}_1^n)\right] \\
&\overset{(b)}{\geq} \mathbf{E}\left[\operatorname{var}(g(X_1^n) \mid \widehat{Y}_j, \{X_i\}_{i \neq j})\right] \\
&\overset{(c)}{\geq} \mathbf{E}\left[\operatorname{var}(g(X_1^n) \mid w_j(X_j), \{X_i\}_{i \neq j})\right] \\
&\overset{(d)}{=} \mathbf{E}\left[\operatorname{var}(g(X_1^n) \mid w_j(X_j), \{X_i\}_{i \neq j}) \mid A\right] \mathbf{P}(A) \\
&\quad + \mathbf{E}\left[\operatorname{var}(g(X_1^n) \mid w_j(X_j), \{X_i\}_{i \neq j}) \mid A^c\right] \mathbf{P}(A^c) \\
&\overset{(e)}{\geq} \mathbf{E}\left[\operatorname{var}(g(X_1^n) \mid w_j(X_j), \{X_i\}_{i \neq j}) \mid A\right] \mathbf{P}(A), \qquad (44)
\end{aligned}$$

where A is the event $X_j \in S \cup t(S)$. Step (a) will hold with equality when the optimal estimate (the conditional expectation of $g(X_1^n)$ given the quantized values) is used; (b) holds because, for each $i \neq j$, $\widehat{Y}_i$ is a function of X_i; (c) holds because $\widehat{Y}_j$ is a function of $w_j(X_j)$; (d) is an application of the total expectation theorem; and (e) holds because the discarded term is nonnegative. It remains to use the hypotheses of the theorem to bound the conditional variance in the final expression.

Since the function is equivalence free, for every $s \in S$,

$$\mathbf{E}\left[\operatorname{var}\left(g(X_1^n) \mid X_j \in \{s, t(s)\}, \{X_i\}_{i \neq j}\right)\right] > 0.$$

Thus

$$\begin{aligned}
\delta &= \int_{s \in S} \mathbf{E}\left[\operatorname{var}\left(g(X_1^n) \mid X_j \in \{s, t(s)\}, \{X_i\}_{i \neq j}\right)\right] f_{X_j}(s)\, ds \\
&> 0.
\end{aligned}$$

Finally, δ is a lower bound to (44) because integrating over $s \in S$ forms the event A and conditioning on $X_j \in \{s, t(s)\}$ is more restrictive than conditioning on the value of $w_j(X_j)$.

Appendix E
Proof of Theorem 11

For brevity, the proof will rely on notation and computations from Appendix B, and details closely paralleling Appendix B are omitted. The basics of using a Taylor expansion with remainder to bound the distortion are unchanged, and the computations up to (41) do not depend on positivity of λ_js. Evaluating (37) using (41) and the negligibility of the Taylor remainder terms gives

$$D \approx \sum_{j=1}^{n} \sum_{i_1^n \in \mathcal{I}_1^n} g_j^2(\beta_{i_1^n}) \mathbf{E}\left[\left(X_j - \beta_{i_j}^{(j)}\right)^2 \mid X_1^n \in S_{i_1^n}\right] \mathbf{P}\left(X_1^n \in S_{i_1^n}\right).$$

In this expression, every term with $\beta_{i_j}^{(j)} \in Z_j$ is zero because $g_j^2(\beta_{i_1^n}) = 0$ by definition of a don't-care interval. Removing the zero terms, estimating partition cell lengths with point densities, and replacing sums with integrals gives

$$D \approx \sum_{j=1}^{n} \frac{1}{12\widetilde{K}_j^2} \int \left(g_j(x_1^n)/\lambda_j(x_j)\right)^2 f_{X_1^n}(x_1^n)\, dx_1^n,$$

where $\widetilde{K}_j$ is the number of cells of Q_j allocated to $[0,1] \setminus Z_j$ and the integration is over $[0,1]^{j-1} \times ([0,1] \setminus Z_j) \times [0,1]^{n-j}$. This can be expressed in a more conceptually transparent way as

$$D \approx \sum_{j=1}^{n} \frac{1}{12\widetilde{K}_j^2} \mathbf{E}\left[\left(\frac{\gamma_j(X_j)}{\lambda_j(X_j)}\right)^2 \mid X_j \notin Z_j\right] \mathbf{P}\left(X_j \notin Z_j\right).$$

The distortion is minimized by making the $\widetilde{K}_j$s as large as possible, which is to set $\widetilde{K}_j = K_j - M_j$, reserving M_j codewords to specify the don't-care intervals.[4] This proves (28) and (29). The optimization of the point densities for fixed-rate quantization follows as in Theorem 8, yielding (30). The final expression (31) follows simply by noting that we are considering the limits as the K_js grow, in which case $2^{R_j} = K_j \approx K_j - M_j$.

Appendix F
Proof of Theorem 12

It is already shown in Theorem 11 that it is optimal to allot a single codeword to each don't-care interval and that the distortion expression (29) then holds. After an appropriate rate analysis, we will optimize the point densities outside of the don't-care intervals.

The key technical problem is that the rate analysis (4) does not hold when there are intervals where f_X is positive but λ is not. This is easily remedied by only applying (4) conditioned on A_j:

$$H(\widehat{X}_j \mid A_j) \approx h(X_j \mid A_j) + \log_2(K_j - M_j) + \mathbf{E}\left[\log_2 \lambda_j(X_j) \mid A_j\right].$$

Now conditioned on A_j, the dependence of distortion and rate on λ_j is precisely in the standard form of Section IV. Thus, following Theorem 8, the optimal point density outside of Z_j is given by (19).

Since the previous results now give the distortion in terms of the conditional entropies $H(\widehat{X}_j \mid A_j)$, what remains is to relate these to the rates:

$$\begin{aligned} R_j &= H(\widehat{X}_j) \\ &\overset{(a)}{=} H(\widehat{X}_j, I_j) \\ &= H(I_j) + H(\widehat{X}_j \mid I_j) \\ &\overset{(b)}{=} H(I_j) + \mathbf{P}\left(A_j\right) H(\widehat{X}_j \mid A_j), \end{aligned}$$

where (a) uses that I_j is a deterministic function of $\widehat{X}_j$; and (b) uses that specifying any $I_j \neq 0$ determines $\widehat{X}_j$ uniquely. Rearranging in anticipation of evaluating (29),

$$\begin{aligned} \log_2(K_j - M_j) \approx\ & \left(\mathbf{P}\left(A_j\right)\right)^{-1} \left(R_j - H(I_j)\right) \\ & - h(X_j \mid A_j) - \mathbf{E}\left[\log_2(\lambda_j(X_j) \mid A_j\right]. \end{aligned}$$

Now evaluating (29) with optimal point densities (19) gives (32).

Appendix G
A Quasi-Triangle Inequality

Lemma 15: Let x and y be functions $\mathbb{R} \to \mathbb{R}^+$ with finite $\mathcal{L}^{1/3}$ pseudonorms. Then

$$\|x\|_{1/3} + \|y\|_{1/3} \geq \frac{1}{4}\|x + y\|_{1/3}.$$

Proof: First, we prove the relation $4(a^3 + b^3) \geq (a+b)^3$ for positive real numbers a and b:

$$\begin{aligned} & 4(a^3 + b^3) - (a+b)^3 \\ &= 4a^3 + 4b^3 - a^3 - b^3 - 3a^2b - 3ab^2 \\ &= 3(a+b)(a-b)^2 \ \geq\ 0. \end{aligned}$$

[4]The argument given presumes minimization for given K_js. But it holds for the variable-rate case as well: there is clearly no benefit to splitting don't-care intervals at a cost of increased rate with no decrease in distortion.

Now by this relation, with $a = \int x(t)^{1/3}\,dt$ and $b = \int y(t)^{1/3}\,dt$:

$$\begin{aligned}
\|x\|_{1/3} + \|y\|_{1/3} &= \left(\int x(t)^{1/3}\,dt\right)^3 + \left(\int y(t)^{1/3}\,dt\right)^3 \\
&\geq \frac{1}{4}\left(\int \left(x(t)^{1/3} + y(t)^{1/3}\right)\,dt\right)^3 \\
&\geq \frac{1}{4}\left(\int \left((x(t) + y(t))^{1/3}\right)\,dt\right)^3 \\
&= \frac{1}{4}\|x+y\|_{1/3},
\end{aligned}$$

where the second inequality uses, pointwise over t, the concavity of the cube-root function on $[0, \infty)$. ■

REFERENCES

[1] V. Doshi, D. Shah, and M. Médard, "Source coding with distortion through graph coloring," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT 2007)*, Jun. 2007, pp. 1501–1505.

[2] R. M. Gray and D. L. Nehoff, "Quantization," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2325–2383, Oct. 1998.

[3] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. IT-19, no. 4, pp. 471–480, Jul. 1973.

[4] S. D. Servetto, "Achievable rates for multiterminal source coding with scalar quantizers," in *Conf. Rec. 39th Asilomar Conf. Signals, Syst. Comput.*, Oct. 2005, pp. 1762–1766.

[5] A. B. Wagner, S. Tavildar, and P. Viswanath, "Rate region of the quadratic Gaussian two-terminal source-coding problem," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 1938–1961, May 2008.

[6] T. S. Han and K. Kobayashi, "A dichotomy of functions $F(X, Y)$ of correlated sources (X, Y) from the viewpoint of the achievable rate region," *IEEE Trans. Inf. Theory*, vol. IT-33, no. 1, pp. 69–76, Jan. 1987.

[7] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. IT-22, no. 1, pp. 1–10, Jan. 1976.

[8] R. Zamir, "The rate loss in the Wyner-Ziv problem," *IEEE Trans. Inf. Theory*, vol. 42, no. 6, pp. 2073–2084, Nov. 1996.

[9] D. Rebollo-Monedero, S. Rane, A. Aaron, and B. Girod, "High-rate quantization and transform coding with side information at the decoder," *Signal Process.*, vol. 86, no. 11, pp. 3160–3179, Nov. 2006.

[10] H. Feng, M. Effros, and S. A. Savari, "Functional source coding for networks with receiver side information," in *Proc. 42nd Annu. Allerton Conf. Commun. Control Comput.*, Sep. 2004, pp. 1419–1427.

[11] E. Martinian, G. Wornell, and R. Zamir, "Source coding with encoder side information," *IEEE Trans. Inf. Theory*, vol. 54, no. 10, pp. 4638–4665, Oct. 2008.

[12] T. Linder, R. Zamir, and K. Zeger, "On source coding with side-information-dependent distortion measures," *IEEE Trans. Inf. Theory*, vol. 46, no. 7, pp. 2697–2704, Nov. 2000.

[13] A. Orlitsky and J. R. Roche, "Coding for computing," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 903–917, Mar. 2001.

[14] V. Doshi, D. Shah, M. Médard, and S. Jaggi, "Distributed functional compression through graph coloring," in *Proc. IEEE Data Compression Conf. (DCC 2007)*, Mar. 2007, pp. 93–102.

[15] H. Yamamoto and K. Itoh, "Source coding theory for multiterminal communication systems with a remote source," *Trans. IECE Japan*, vol. E63, no. 10, pp. 700–706, Oct. 1980.

[16] J. A. Bucklew, "Multidimensional digitization of data followed by a mapping," *IEEE Trans. Inf. Theory*, vol. IT-30, no. 1, pp. 107–110, Jan. 1984.

[17] T. Linder, R. Zamir, and K. Zeger, "High-resolution source coding for non-difference distortion measures: Multidimensional companding," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 548–561, Mar. 1999.

[18] S. A. Kassam, "Optimum quantization for signal detection," *IEEE Trans. Commun.*, vol. COM-25, no. 5, pp. 479–484, May 1977.

[19] B. Picinbono and P. Duvaut, "Optimum quantization for detection," *IEEE Trans. Commun.*, vol. 36, no. 11, pp. 1254–1258, Nov. 1988.

[20] H. Xie and A. Ortega, "Entropy- and complexity-constrained classified quantizer design for distributed image classification," in *Proc. 2002 IEEE Workshop Multimedia Signal Process.*, Dec. 2002, pp. 77–80.

[21] L. Vasudevan, A. Ortega, and U. Mitra, "Application-specific compression for time delay estimation in sensor networks," in *Proc. 1st Int. Conf. Embedded Netw. Sensor Syst. (SenSys'03)*, Nov. 2003, pp. 243–254.

[22] T. S. Han and S.-I. Amari, "Statistical inference under multiterminal data compression," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2300–2324, Oct. 1998.

[23] J. Li, N. Chaddha, and R. M. Gray, "Asymptotic performance of vector quantizers with a perceptual distortion measure," *IEEE Trans. Inf. Theory*, vol. 45, no. 4, pp. 1082–1091, May 1999.

[24] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Boston: Kluwer Academic Publishers, 1992.

[25] J. J. Y. Huang and P. M. Schultheiss, "Block quantization of correlated Gaussian random variables," *IEEE Trans. Commun. Syst.*, vol. CS-11, no. 3, pp. 289–296, Sep. 1963.

[26] B. Farber and K. Zeger, "Quantization of multiple sources using nonnegative integer bit allocation," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 4945–4964, Nov. 2006.

[27] H. Viswanathan and R. Zamir, "On the whiteness of high-resolution quantization errors," *IEEE Trans. Inf. Theory*, vol. 47, no. 5, pp. 2029–2038, Jul. 2001.

[28] M. Studený and J. Vejnarová, "The multiinformation function as a tool for measuring stochastic dependence," in *Learning in Graphical Models*, M. I. Jordan, Ed. Dodrecht: Kluwer Academic Publishers, 1998, pp. 261–297.

[29] V. Misra, "Functional quantization," Master's thesis, Massachusetts Institute of Technology, Jun. 2008.

[30] A. György, T. Linder, and K. Zeger, "On the rate–distortion function of random vectors and stationary sources with mixed distributions," *IEEE Trans. Inf. Theory*, vol. 45, no. 6, pp. 2110–2115, Sep. 1999.